What Is Probability?

Some mathematicians argue that it is "statistical"; others, that it is "inductive." The author believes that there are two kinds, both essential to the future progress of science

by Rudolf Carnap

The articles in this issue on fundamental questions of science give an illuminating picture of the way scientists work. No one reading these articles can fail to be impressed with the great importance to science of hypotheses—the daring guesses on slender evidence that go into building new theories. The question I should like to raise in this final article is: Can the method of scientific inquiry be made more precise? Can we learn to judge the hypotheses, to weigh the extent to which they are supported by the evidence at hand, as an investigator judges and weighs his data?

The question leads at once into the subject of probability. If you query scientists about the meaning of this term, you will discover a curious situation. Practically everyone will say that probability as used in science has only one meaning, but when you ask what that meaning is, you will get different answers. Most scientists will define it as statistical probability, which means the relative frequency of a given kind of events or phenomena within a class of phenomena, usually called the "population." For instance, when a statistician says the probability that a native of the U. S. has A-type blood is 4/10, he means that four out of 10 people have this type. This meaning of probability has become almost the standard usage in science. But you will also find that there are scientists who define probability in another way. They prefer to use the term in the sense nearer to everyday use, in which it means a measurement, based on the available evidence, of the chances that something is true—as when a jury decides that a defendant is "probably" guilty, or a weather forecaster predicts that it will probably rain tomorrow. This kind of probability amounts to a weighing of the strength of the evidence. Its numerical expression has a meaning quite different from that of statistical probability: if the weather man were to venture to say that the probability of rain tomorrow was 4/10, he would not be describing a statistical fact but would simply mean that, should you bet on it raining tomorrow, you had better ask for odds of 4 to 6.

This concept is called inductive probability. A scientist makes a judgment of the odds consciously or unconsciously, whenever he plans an experiment. Usually the probability ascribed to his hypothesis is stated not in numbers but in comparative terms; that is, the probability is said to be high or low, or one probability is considered higher than another. To some of us it seems that inductive probability could be refined into a more precise tool for science. Given a hypothesis and certain evidence, it is possible to determine, by logical analysis and mathematical calculation, the probability that the hypothesis is correct, or the "degree of confirmation." If we had a system of inductive logic in mathematical form, our inferences about hypotheses in science, business and everyday life, which we usually make by "intuition" or "instinct," might be made more rational and exact. I have made a beginning in the construction of such a system, using the findings of past workers in this field and the exact tools of modern symbolic logic. Before discussing this system, let me review briefly the history of the inductive concept of probability.

The scientific theory of probability began, as a matter of fact, with the inductive concept and not the statistical one. Its study was started in the 16th century by certain mathematicians who were asked by their gambler friends to determine the odds in various games of chance. The first major treatise on probability, written by the Swiss professor Jacob Bernoulli and published posthumously in 1713, was called *Ars Conjectandi*, "The Art of Conjecture"—in other words, the art of judging hypotheses on the basis of evidence. The classical period in the study of probability culminated in the great 1812 work *Théorie analytique des probabilités*, by the French astronomer and mathematician Pierre Laplace. He declared the aim of the theory of probability to be to guide our judgments and to protect us from illusions, and he was concerned primarily not with statistics but with methods for weighing the acceptability of assumptions.

But after the middle of the 19th century the word probability began to acquire a new meaning, and scientists turned more and more to the statistical concept. By the 1920s Robert Aylmer Fisher in England, Richard von Mises and Hans Reichenbach in Germany (both of whom have died within the last few months) and others began to develop new probability theories based on the statistical interpretation. They were able to use many of the mathematical theorems of classical probability, which hold equally well in statistical probability. But they had to reject some. One of the principles they rejected, called the principle of indifference, sharply points up the distinction between inductive and statistical probability.

Suppose you are shown a die and are told merely that it is a regular cube. With no more information than this, you can only assume that when the die is thrown any one of its six faces is as likely to turn up as any other; in other words, that each face has the same probability, 1/6. This illustrates the principle of indifference, which says that if the evidence does not contain anything that would favor one possible event over another, the events have equal probabilities *relative to this evidence*. Now a second observer may have additional evidence: he

	STATISTICAL DISTRIBUTIONS: NUMBER OF BLUE	NUMBER OF WHITE	INDIVIDUAL DISTRIBUTIONS		METHOD I: INITIAL PROBABILITY OF INDIVIDUAL DISTRIBUTIONS	METHOD II: INITIAL PROBABILITY OF: STATISTICAL DISTRIBUTIONS	INDIVIDUAL DISTRIBUTIONS
1.	4	0	1.	● ● ● ●	1/16	1/5	1/5 = 12/60
			2.	● ● ● ○	1/16		1/20 = 3/60
2.	3	1	3.	● ● ○ ●	1/16	1/5	1/20 = 3/60
			4.	● ○ ● ●	1/16		1/20 = 3/60
			5.	○ ● ● ●	1/16		1/20 = 3/60
			6.	● ● ○ ○	1/16		1/30 = 2/60
			7.	● ○ ● ○	1/16		1/30 = 2/60
3.	2	2	8.	● ○ ○ ●	1/16	1/5	1/30 = 2/60
			9.	○ ● ● ○	1/16		1/30 = 2/60
			10.	○ ● ○ ●	1/16		1/30 = 2/60
			11.	○ ○ ● ●	1/16		1/30 = 2/60
			12.	● ○ ○ ○	1/16		1/20 = 3/60
4.	1	3	13.	○ ● ○ ○	1/16	1/5	1/20 = 3/60
			14.	○ ○ ● ○	1/16		1/20 = 3/60
			15.	○ ○ ○ ●	1/16		1/20 = 3/60
5.	0	4	16.	○ ○ ○ ○	1/16	1/5	1/5 = 12/60

INDUCTIVE PROBABILITY METHODS are illustrated in an example which is tabulated above. Four balls are to be drawn in succession from an urn. They are identical in every way except that some are blue and some white. Nothing is known, however, about the proportion of blue to white balls in the urn. First we want to decide on the initial probabilities in the experiment—the probabilities before the first ball is drawn. We list (under "Individual Distributions") all the possible ways in which the drawing can turn out. Now we apply the principle of indifference, which says that if the evidence contains nothing that favors one possibility over another, all possibilities must be considered equally probable. There are two ways to apply the principle to this example. The first is illustrated under "Method I." Since there are 16 possible cases, dividing the probabilfty equally among them gives each a probability of 1/16. But there is another way to look at the table. Instead of taking into account the order in which blue and white turn up, we can concentrate only on the total numbers of blue and white in a drawing—all blue, three blue and one white, and so on. This classifies the table into "Statistical Distributions," which are indicated by the brackets on the left. There are five statistical distributions. If the principle of indifference is applied to them, then each has a probability of 1/5, as shown in the first column of "Method II." Now the individual distributions within each statistical distribution are assigned probabilities that are again determined by the principle of indifference. The first statistical distribution (four blue) has only one member, so it gets the full amount of the probability to be distributed, or 1/5, as shown in the second column of "Method II." The second statistical distribution (three blue, one white) has four members, so the probability must be split four ways, 1/20 to each. Similarly, the remaining three statistical distributions are divided into their individual members. At the extreme right hand of the table, all probabilities are converted to a least common denominator of 60 in order to facilitate comparing and combining them. Method II is superior to Method I because it assigns probabilities to future events on the basis of the frequency of their past occurrence.

may know that the die is loaded in favor of one of the faces, without knowing which face it is. The probabilities are still the same for him, because as far as his information goes, each of the six faces has an equal possibility of being loaded. On the other hand, for a third observer who knows that the load favors the face numbered 1 the probabilities change; on the basis of his evidence the probability of the ace is higher than 1/6.

Thus inductive probability depends on the observer and the evidence in his possession; it is not simply a property of the object itself. In statistical probability, which refers to the actual frequency of an event, the principle of indifference is of course absurd. It would be incautious for an observer who knew only that a die had the accurate dimensions of a cube to assert that the six faces would appear with equal frequency. And if he knew that the die was biased in favor of one side, he would contradict his own knowledge. Inductive probability, on the other hand, does not predict frequencies; rather, it is a tool for evaluating evidence in relation to a hypothesis. Both the statistical and inductive concepts of probability are indispensable to science; each has valuable functions to perform. But it is important to recognize the distinctions between the two concepts and to develop the possibilities of both tools.

In the past 30 years the inductive concept of probability, which had been supplanted by the statistical concept, has been revived by a few workers. The first of these was the great English economist John Maynard Keynes. In his *Treatise on Probability* in 1921 he showed how the inductive concept is implicitly used in all our thinking about unknown events, in science as well as in everyday life. Yet Keynes' attempt to develop this concept was too restricted: he believed it was impossible to calculate numerical probabilities except in well-defined situations such as the throw of dice, the possible distributions of cards, and so on. Moreover, he rejected the statistical concept of probability and argued that all probability statements could be formulated in terms of inductive probability.

I believe that he was mistaken in this point of view. Today an increasing number of those who study both sides of the controversy, which has been going on for 30 years, are coming to the conclusion that here, as often before in the history of scientific thinking, both sides are right in their positive theses, wrong in their polemical remarks. The statistical concept, for which a very elaborate mathematical theory exists, and which has been applied fruitfully in many fields in science and industry, need not be abandoned in order to make room for the inductive concept. Statistical probability characterizes an objective situation, *e.g.*, a state of a physical, biological or social system. On the other hand, inductive probability, as I see it, does not occur in scientific statements but only in judgments *about* such statements. Thus it is applied in the methodology of science—the analysis of concepts, statements and theories.

In 1939 the British geophysicist Harold Jeffreys put forward a much more comprehensive theory of inductive probability than Keynes'. He agreed with the classical view that probability can be expressed numerically in all cases. Furthermore, he wished to apply probability to quantitative hypotheses of science, and he set up an axiom system for probability much stronger than that of Keynes. He revived the principle of indifference in a form which seems to me much too strong: "If there is no reason to believe one hypothesis rather than another, the probabilities are equal." It can easily be shown that this statement leads to contradictions. Suppose, for example, that we have an urn known to be filled with blue, red and yellow balls but do not know the proportion of each color. Let us consider as a starting hypothesis that the first ball we draw from the urn will be blue. According to Jeffreys' (and Laplace's) statement of the principle of indifference, if the question is whether the first ball will be blue or not blue, we must assign equal probabilities to both these hypotheses; that is, each probability is 1/2. If the first ball is not blue, it may be either red or yellow, and again, in the absence of knowledge about the actual proportions in the urn, these two have equal probabilities, so that the probability of each is 1/4. But if we were to start with the hypothesis that the first ball drawn would be, say, red, we would get a probability of 1/2 for red. Thus Jeffreys' system as it stands is inconsistent.

In addition, Jeffreys joined Keynes in rejecting the statistical concept of probability. Nevertheless his book *Theory of Probability* remains valuable for the new light it throws on many statistical problems by discussing them for the first time in terms of inductive probability.

I have drawn upon the work of Keynes and Jeffreys in constructing my mathematical theory of inductive probability, set forth in the book *Logical Foundations of Probability*, which was published in 1950. It is not possible to outline here the mathematical system itself. But I shall explain some of the general problems that had to be solved and some of the basic conceptions underlying the construction.

One of the fundamental questions to be decided is whether to accept a principle of indifference, and if so, in what form. It should be strong enough to allow the derivation of the desired theorems, but at the same time sufficiently restricted to avoid the contradictions resulting from the classical form.

The problem can be made clear by an example illustrating a few elementary concepts of inductive logic. We have an urn filled with blue and white balls in unknown proportions. We are going to draw four balls in succession. Taking the order into account, there are 16 possible drawings (all four blue, the first three blue and the fourth white, the first white and the next three blue, and so on). We list these possibilities in a table (*see table on page 129*).

Now what is the initial probability, before we have drawn at all, that we shall draw any one of these 16 distributions? We might assign any probability to the individual distributions, so long as they all added up to 1. Suppose we apply the principle of indifference and say that all the distributions have equal probabilities; that is, each has a probability of 1/16.

Let us state a specific hypothesis and calculate its probability. The hypothesis is, for example, that among the first three balls we draw, just one will be white. Looking at the table, we can see that six out of the 16 possible drawings will give us this result. The probability of our hypothesis, therefore, is the sum of these initial probabilities, or 6/16.

Suppose now that we are given some evidence, *i.e.*, have drawn some balls, and are asked to calculate the probability of a given hypothesis on the basis of this evidence. For instance, we have drawn first a blue ball, then a white ball, then a blue ball. The hypothesis is that the fourth ball will be blue; what is its probability? Here we run into a question as to how we should apply the principle of indifference. Let us try two different methods.

In Method I we start by assigning equal probabilities to the individual distributions. Referring to the table, we see that two of these distributions (Nos. 4 and 7) will give us the sequence blue, white, blue for the first three balls. Its probability is therefore 2/16. In only one of these distributions is the fourth ball blue; its probability is 1/16. The probability of our hypothesis on the basis of the evidence is obtained by dividing one into the other: *i.e.*, 1/16 divided by 2/16, which equals 1/2. In other words, the chances that our hypothesis is correct are 50-50: the fourth ball is just as likely to be white as blue.

But as a guide to judging a hypothesis, this result contradicts the principle of learning from experience. Other things being equal, we should consider one event more probable than another if it has happened more frequently in the past. We would regard a man as unreasonable if his expectation of a future event were the higher the less often he had seen it before. We must be guided by our knowledge of observed events, and in this example the fact that two out of three balls drawn from an unknown urn were blue should lead us to expect the probabilities to favor the fourth's also being blue. Yet a number of philosophers, including Keynes, have proposed Method I in spite of its logical flaw.

There is a second method which gives us a more reasonable result. We first apply the principle of indifference not to individual distributions but to statistical distributions. That is, we consider only the number of blue balls and of white balls obtained in a drawing, irrespective of order. The table shows that there are five possible statistical distributions (four blue, four white, three blue and one white, three white and one blue, two blue and two white). By the principle of indifference we assign equal probabilities to these, so that the probability of each is 1/5. We distribute this value (expressed for arithmetical convenience as 12/60) in equal parts among the corresponding individual distributions (*see last column of table*). Now the probabilities of distributions No. 4 and No. 7 are 3/60 and 2/60, respectively, and the probability of the hypothesis on the basis of the evidence is 3/60 divided by 5/60, or 3/5. In short, the chances that the fourth ball will be blue are not even but 3 to 2, which is more consistent with what experience, meaning the evidence we have acquired, should lead us to expect.

Method II, as well as Method I, leads to contradictions if it is applied in an unrestricted way. If it is used in cases characterized by more than one property difference (such as the difference between blue and white balls in our example) then all the relevant differences must be specified. Thus restricted, this system, which I proposed in 1945, is the first consistent inductive method, so far as I am aware, that succeeded in satisfying the principle of learning from experience. Since then I have found that there are many others. None of them seems as simple to define as Method II, but some of them have other advantages.

Having found a consistent and suitable inductive method, we can proceed to develop a general procedure for calculating, on the basis of given evidence, an estimate of an unknown value of any quantity. Suppose that the evidence indicates a certain number of possible values for a quantity at a given time, *e.g.*, the amount of rain tomorrow, the number of persons coming to a meeting, the price of wheat after the next harvest. Let the possible values be x_1, x_2, x_3, etc., and their inductive probabilities be p_1, p_2, p_3, etc. Then p_1x_1 is the "expectation value" of the first case at the present moment, p_2x_2 of the second case, and so on. The total expectation value of the quantity on the given evidence is the sum of the expectation values for all the possible cases. To take a specific example, suppose there are four prizes in a lottery, a first prize of $200 and three prizes of $50 each. It is known that the probability of a ticket winning the first prize is 1/100, and of a second prize, 3/100; the probability that the ticket will win nothing is therefore 96/100. Applying the method I have described above, a ticket holder can estimate that the ticket is worth to him 1/100 times $200 plus 3/100 times $50 plus 96/100 times 0, or $3.50. It would be irrational to pay more for it.

The same method may be used to make a rational decision in a situation where one among various possible actions is to be chosen. For example, a man considers several possible ways of investing a certain amount of money. He can—in principle, at least—calculate the estimate of his gain for each possible way. To act rationally, he should then choose that way for which the estimated gain is highest.

Bernoulli, Laplace and their followers envisaged a theory of inductive probability which, when fully developed, would supply the means for evaluating the acceptability of hypothetical assumptions in any field of theoretical research and for making rational decisions in the affairs of practical life. They were a great deal farther from this audacious objective than they realized. In the more sober cultural atmosphere of the late 19th and early 20th centuries their idea was dismissed as Utopian. But today a few men dare to think that these pioneers were not mere dreamers.