20 February 1959, Volume 129, Number 3347 SCIENCE

Kinds of Probability

Although there are at least five kinds of probability, we can get along with just one kind.

I. J. Good

The mathematician, the statistician, and the philosopher do different things with a theory of probability. The mathematician develops its formal consequences, the statistician applies the work of the mathematician, and the philosopher describes in general terms what this application consists in. The mathematician develops symbolic tools without worrying overmuch what the tools are for; the statistician uses them; the philosopher talks about them. Each does his job better if he knows something about the work of the other two.

What is it about probability that has interested philosophers? Principally, it is the question whether probability can be defined in terms of something other than itself, and, if not, how the idea of probability is used, what is its meaning, what are the shades of meaning. Can we verify that probability exists, or must we be satisfied to say how it is used? Is the "use" theory of meaning more appropriate than the "verification" theory? It seems to me that the philosopher's job is mainly to describe what a man does or thinks at the precise moment that he uses the idea of probability.

Our main question is this: are there different kinds of probability? The question is analogous to the one "Are there different kinds of life?" In a sense there are two kinds of life: animal and vegetable; in another sense there are as many as there are genera or species; in yet another sense there is only one kind of life, since life is indivisible, and even the distinction between animals and vegetables is misleading in some contexts. (As a matter of fact, even the distinction between living and inanimate matter can mislead people into supposing that evolution is impossible.) Much of the controversy about the theory of probability is like this. From some points of view there are at least five kinds of probability; from another point of view they can all be defined in terms of a single kind. I shall elaborate this remark and begin by describing some different kinds of probability. Classification of different kinds of probability is half the problem of the philosophy of probability.

The author is at the Admiralty Research Laboratory, Teddington, Middlesex, England.

The Classical Definition

Some billion years ago, an anonymous speck of protoplasm protruded the first primitive pseudopodium into the primeval slime, and perhaps the first state of uncertainty occurred. Thousands of years ago words such as *maybe, chance, luck,* and *fate* were introduced into languages. If a theory is a method of using language, we could say that theories of probability are thousands of years old. But often a usage of language is not dignified by the name *theory* unless a real effort has been made to describe this usage accurately: a theory, then, is not just talk, but is also talk about talk. (Philosophers of science talk about talk about talk.) So when Aristotle (about 300 B.C.) said "the probable is what usually happens," and when Cicero (about 60 B.C.) described probability as the "guide of life," they had formulated primitive theories of probability and of rational behavior. We can hardly tell whether these theories had any practical results; at any rate, the ancient Romans later practised insurance, and Domitius Ulpianus drew up a table of life expectancies (about A.D. 200.)

Mathematical ideas, however, date back only a few hundred years. A commentary in 1477 on Dante's *Purgatorio* gives the probabilities of various totals when three dice are thrown. Perhaps the application was to cleromancy (divination by dice). In the 16th century, Cardan, an inveterate gambler, made several simple probability calculations of use to gamblers. He defined probability as a "proportion of equally probable cases"; for example, of the 36 possible results of throwing two dice, three give a total of 11 or more points, so the probability of this event is defined as 1/12 if the 36 possible results are equally probable. The definition by equally probable cases is usually called the "classical definition."

The origin of the mathematical theory of probability is not usually ascribed to Cardan, but rather to Pascal (1654), who, in correspondence with Fermat, solved the first mathematically nontrivial problems. The first book on the subject, of any depth, was published soon afterwards by Huygens.

All of these authors were concerned with games of chance, and although they defined probability as a proportion of equally probable cases, their purpose must have been to explain why certain long-run proportional frequencies of success occurred. Without being explicit about it they were trying to explain one kind of probability in terms of another kind. James Bernouilli was much more explicit about it, in his famous work *Ars Conjectandi,* published in 1713, eight years after his death. His "law of large numbers" states that in n "trials," each with probability p of success, the number of successes will very probably be close to pn if n is large. For example, if a coin has probability exactly ½ of coming down heads, and if it is tossed a thousand times, then the number of heads is unlikely to differ much from 500; more precisely, the proportion of heads is unlikely to differ much from ½. In fact, the number of heads will very probably lie between 470 and 530. In a million tosses the number of heads will very probably lie between 499,000 and 501,000. These results are based on the assumption that the probability of heads is ½ at each throw, no matter what the results of previous throws may have been. In other words, the trials must be "causally independent." Bernouilli did not make it clear that the trials must be causally independent and that pn must be large. If your probability, p, of winning a sweepstake is 1/1,000,000, then Bernouilli's theorem would not be applicable until you had entered several million sweepstakes (by which time you would be too old to care).

Bernouilli proved his theorem on the assumption that the probability, p, was defined as a proportion of equally probable cases. But he tried to apply the theorem to social affairs in which this definition is hardly appropriate. Worse yet: the probability is likely to be variable.

Subjective Probability

Even in games of chance the classical definition is not entirely satisfactory, for the games may not be "fair." A fair game of chance is one in which the apparently equal probabilities "really are" equal. In order to give this definition of a fair game any substance we must again distinguish between two kinds of probability. Consider, for example, the probability that cutting an ordinary pack of playing cards will put a red card at the bottom of the pack, an event that I shall call a "success." Since half the cards are red and half are black, the probability would seem to be ½ if the pack of cards, and its shuffling, are fair. But if all the red cards have dirty, sticky faces, then a black card is more likely to be brought to the bottom. If we knew the red cards had sticky faces we would prefer to bet on a black card, in a "level bet." But if we did not know it, then the probability would still be ½ *for us.* Even if we allowed for the possibility of stickiness, the black cards are as likely to be more sticky as to be less so, unless we have some further information. For us the *first* cut has probability ½ of being successful.

We may have an opponent who knows that the red cards are stickier. For him the probability is not the same as it is for us. This example shows that personal, or subjective, or logical probability depends on the given information as well as on the event whose probability is to be estimated. This is the reason for notation of the form

$$P(E|F)$$

read from left to right (like all good notations) "the probability of E given F." For the sake of generality, E and F may be interpreted as propositions. This notation (or equivalent ones) has become standard during the present century. In this notation the probabilities we have just been discussing are

P(bottom card is red|
the cards have been well shuffled)

and

P(bottom card is red|the cards have
been well shuffled by normal standards,
but the red ones have sticky faces).

The use of the vertical stroke, or equivalent notation, is likely to save us from the errors that may arise through talking simply about the "probability that the bottom card will be red," without reference to the "given" (= assumed) information.

Physical Probability

Suppose that our opponent has carried out a very extensive experiment and has decided that the long-run proportion of successes is 0.47 (instead of ½). We may be tempted to call this the "true probability," or "physical probability," or "material probability," or "chance," or "propensity," and to regard it as having an impersonal, public, or objective significance. Whether or not physical probability is regarded as distinct from personal, private, intuitive, subjective, or logical probability, it is often convenient to talk as if it were distinct. I shall, however, argue later on that its numerical value can be defined in terms of subjective probability.

A physical probability is the probability of a "success" given the "experimental setup." So for physical probabilities, too, it is convenient to have a notation of the form

$$P(E|F)$$

We can distinguish between true and hypothetical probabilities, depending on whether the experimental setup is true or hypothetical. For example, we can take an actual pack of cards and we can discuss the probability that the bottom card will be red "given" (= on the assumption that) all the clubs have been omitted. This probability makes sense even if the clubs have not in fact been omitted, and the probability will then be "hypothetical" and not "true." It so happens that it is decidedly useful to talk about hypothetical probabilities as well as true ones.

We could *imagine* a physical chemist who could analyze the chemicals on the faces of the cards and then compute the probability of success by quantum theory. But this would be a far cry from the simple physical symmetry that led Cardan and Pascal to judgments of equal probability, or from the logical symmetry that caused us to consider black and red to be equally likely to be the stickier. It is perhaps clear by now that the classical definition, however suggestive, is by no means general enough to cover all the uses of the word *probability.*

Inverse Probability

Most applications of the theory of probability to the social sciences are more like unfair games of chance than fair ones. If n smokers are sent questionnaires and r of them refuse to fill them out, what is the probability, p, that the next smoker selected will refuse to fill out his questionnaire? And what is the proportion of all smokers who will refuse? Whereas Bernouilli's theorem works from a knowledge of p to information about the number of "successes" in the sample, the answer here seems to require the inverse process. A simple estimate of p is r/n, but if r is small this may be a poor estimate, especially if $r = 0$. (To say that the probability of an event is 0 is to say that the event is infinitely unlikely. Such an assertion is not justified merely by 100 percent failures in the past.) Sometimes r/n is taken as a definition of probability; it may be called the "naive" definition.

A better attempt at "inverting Bernouilli's theorem" was made by Thomas Bayes in a paper published posthumously in 1763. The method is known as "inverse probability," and was given a prominent place in Laplace's *Théorie analytique des probabilités* (1812). It may also be described as the Bayes-Laplace method of statistical inference. In modern terminology the principle of inverse probability can be expressed in terms of "initial" probabilities, "final" probabilities, and "likelihoods." The initial probability (also called the "prior" probability) of a hypothesis is its probability before some experiment is performed. (There may or may not have been previous experiments or evidence, so the description "a priori" is inappropriate.) The final probability is the probability after the experiment is performed. These probabilities are different, in general, because the given information is different. The *likelihood* of a hypothesis is the probability, given that hypothesis, of the actual result of the experiment.

For example, suppose we have two hypotheses about a coin, either that the coin is fair or that it is double-headed, and suppose that the initial probabilities of these two hypotheses are equal, that is, each is ½. Suppose now that the coin is tossed ten times and comes down heads every time. The likelihoods of the two hypotheses are then $2^{-10} = 1/1024$ and 1.

Bayes' theorem is, in effect, that the final probability of a hypothesis is proportional to its initial probability times its likelihood. In our example, the final probabilities are therefore proportional to 1/1024 and 1. Therefore, the final probability that the coin is double-headed is 1024/1025, or nearly certain.

Although Laplace's exposition was clearer than Bayes', he blatantly assumed that initial probabilities were always equal, whereas Bayes was more modest. Laplace assumed, for example, that an unknown physical probability, p, was initially (that is, before any observations were taken) equally likely to "take any value" between 0 and 1; he assumed, for example, that each of the intervals (0, 0.01), (0.01, 0.02), . . . , (0.99, 1.00) initially had probability 0.01. In the applications, p is what we are calling a physical probability existing "out there," whereas the probability 0.01 is a more subjective kind of probability. By making this assumption of a "uniform distribution" of probability between 0 and 1, Laplace proved his so-called "law of succession." This states that after r "successes" in n "trials," p can be estimated as

$$(r+1)/(n+2)$$

For example, after one success in two trials, p is estimated as ½; after one success in one trial, p is estimated as 2/3; after no success in one trial, p is estimated as 1/3; after no success in no trials, p is estimated as ½. The formula is open to dispute and has often been disputed. It leads, for example, to the conclusion that anything that has been going on for a given length of time has probability ½ of going on for the same length of time again. This does not seem to me to be too bad a rule of thumb if it is applied with common sense.

Inverse probability is by no means the only method of statistical inference. There is, for example, an important method known as "maximum likelihood," used at times by Daniel Bernouilli (1777), Gauss (1823), and especially by Fisher (1912). In this method, that hypothesis is selected whose likelihood is a maximum, where "likelihood" is defined as it is above. For the simple sampling experiment mentioned above, the method of maximum likelihood leads to the naive estimate r/n, which in my opinion is not as good as the result given by Laplace's law of succession.

A familiar objection to the use of inverse probability is that the initial probabilities cannot usually be determined by clear-cut rules. The method of maximum likelihood is clear-cut, and does not lend itself so easily to conscious or unconscious cheating. But for small samples it can lead to absurd conclusions. The method of inverse probability, although more arbitrary, need never lead to absurdity unless it is dogmatically combined with an assumption that the initial probabilities of alternative hypotheses are invariably equal.

Definition by Long-run Frequency

One of Laplace's tricks was to use the expression "equally possible cases" instead of "equally probable cases," and thereby to pretend that he had defined probability completely. Not many people today are taken in by this verbal trick.

Leslie Ellis (1843), A. Cournot (1843), G. Boole (1854), and J. Venn (in a full-length treatise, 1866), were not taken in. They asked, for example, how you could prove that a die was unloaded except by throwing it a great number of times. They proposed to solve the problem of inverting Bernouilli's theorem by simply defining physical probability in terms of long-run frequency ("frequentism").

If a roulette wheel is spun 300 times and there is no occurrence of a 7 should we regard the probability of a 7 on the next spin as 1/37 (its "official" value), or as 0, or as some intermediate value? This simple question exposes the weakness both of Laplace's position and of pure frequentism. The frequentist would perhaps refuse to make any estimate and would say "spin the wheel another few hundred times." Owing to lack of space I shall leave this question and consider an even simpler one.

Suppose that a coin-spinning machine is set to work and produces the sequence

HTHTHTHTHTHTHTHTHTHT

The proportion of heads is precisely ½ and it seems reasonable to predict that the "Venn limit," that is, the limiting proportion of heads if the sequence is indefinitely continued, will also be ½. Yet no one would say that the spinning was fair. This type of difficulty was recognized by Venn but was not adequately met. R. von Mises (1919) proposed a new frequentist theory of probability based on the notion of infinitely long random sequences—what he called "irregular collectives." The main property of an irregular collective is that the proportion of "successes" (say heads) is the same for every sub-sequence selected in advance. This property is closely related to the impossibility of a successful gambling system. An irregular collective is an abstraction like a point in Euclidean geometry. Von Mises drew a clear distinction between the mathematical or abstract theory and the problem of application of that theory. He was perhaps the first person to make this distinction explicit for the theory of probability, in other words, to advocate Euclid's method, the "axiomatic method." But having made the distinction, he virtually ignored the philosophical problem of application. He stated, like the 19th-century frequentists, that in the applications the sequences must be long, but he did not say how long; just as the geometer might say that dots must be small before they are called points, without saying how small. But the modern statistician often uses small samples; he is like a draftsman with a blunt pencil. He would like to know how long is a long run. As J. M. Keynes said, "In the long run we shall all be dead."

If a frequentist is cross-examined about how long is a long run, it is possible to deduce something about the implicit initial probabilities that he is using. This can be done algebraically, by assuming that the initial probabilities exist as "unknowns," applying the theory of probability, including Bayes' theorem, making use of the frequentist's judgments, and finally solving for the unknown initial probabilities (or getting upper and lower bounds for them). In this way the frequentist may be seen, in spite of hot denials, to be behaving *as if* he had judgments concerning initial probabilities of hypotheses. Or he may be caught in a contradiction.

Like Venn, von Mises deliberately restricted the generality of the theory to situations where the long-run frequency definition seemed to be reasonable. He was entitled to do this but he was not justified in being intolerant of theories that try to achieve more, and especially those that concern themselves more with the philosophical problem of applicability.

Among other brilliant mathematicians since von Mises who have developed the mathematical theory, perhaps Kolmogorov deserves special mention. Most of these mathematicians have been concerned both with the mathematical theory and with its applications, but much less with the philosophical problem of applicability. Among those who have been so concerned were the philosopher W. E. Johnson, his two pupils J. M. Keynes (*1*) and H. Jeffreys (*2*), F. P. Ramsey (*3*), B. de Finetti (*4*), B. O. Koopman (*5*), R. Carnap (*6*), B. Russell (*7*), I. J. Good (*8*), and L. J. Savage (*9*).

Neoclassical Definition

Some of these writers are dualists and hold that one should talk about two kinds of probability. Others put most emphasis on the subjectivistic or logical interpretation. Here I shall merely summarize some of my own views, which in one respect or another are closely related to those of the other authors just mentioned. The theory may reasonably be called "neoclassical" or "neo-Bayesian," since its opponents are primarily frequentists, and since Bayes' theorem is restored to a primary position from which it had been deposed by the orthodox statisticians of the second quarter of the 20th century, especially by R. A. Fisher.

1) The function of the theory of subjective probability is to introduce as much objectivity (impersonality) as possible into "your" subjective body of beliefs, not to make it completely impersonal, which may be impossible. With the help of a mathematical theory, based on a few axioms, a body of beliefs can be enlarged and inconsistencies in it can be detected. A subjective probability is a degree of belief that belongs to a body of beliefs from which the worst inconsistencies have been removed by means of detached judgments.

2) Subjective probabilities are not usually precise but are circumscribed by inequalities ("taking inequalities seriously" or "living with vagueness").

3) Probability judgments are plugged into a sort of black box (the abstract or mathematical theory) and discernments are fed out; the judgments can be of very varied type, so that nothing of value in frequentism, classicism, or any other theory is lost.

4) Many orthodox statistical techniques achieve objectivity only by throwing away information, sometimes too much. One way this can happen is if the observations supplied by a very expensive experiment support a hypothesis not thought of in advance of the experiment. In such circumstances, it will often happen that the experimenter will be thrown back on his personal judgment.

5) The theory can be extended to become a theory of rational behavior, by introducing "utilities" (value judgments).

6) All this is important for statistical practice and for the making of decisions.

7) A theory of subjective probability is general enough to cover physical probabilities, but not conversely. Although a physical probability can be regarded as something that is not subjective, its numerical value can be equated to the limiting value of a subjective probability when an experiment is repeated indefinitely under essentially constant circumstances.

Kinds of Probability

Since this article is concerned mainly with subjective and physical probability, it would be inappropriate to discuss other kinds in great detail. Perhaps a mere list of various kinds will be of interest:

1) Degree of belief (intensity of conviction), belonging to a highly self-contradictory body of beliefs. (This hardly deserves to be called a probability.)

2) Subjective probability (personal probability, intuitive probability, credence). Here some degree of consistency is required in the body of beliefs.

3) Multisubjective probability (multicredence). The name here is self-explanatory.

4) Credibility (logical probability; impersonal, objective, or legitimate intensity of conviction).

5) Physical probability (material probability, chance, propensity; this last name was suggested by K. R. Popper).

6) Tautological probability. In modern statistics it is customary to talk about ideal propositions known as "simple statistical hypotheses." If, for each possible result, E, of an experiment, $P(E|H)$ is equal to a number that is specified as part of the *definition* of H, then the probability $P(E|H)$ may be called a "tautological probability," and H is a "simple statistical hypothesis."

Much of statistics is concerned with testing whether a simple statistical hypothesis is "true" (or approximately true) by means of sampling experiments. If we regard this as more than a manner of speaking, then, for consistency, we must believe in the existence of physical probabilities. For example, the proposition that a coin is unbiased is a simple statistical hypothesis, H, part of whose *definition* is that $P(\text{heads}|H) = 1/2$, a tautological probability. But if we say or believe that this proposition is *true,* then we are committed to saying or believing also that this tautological probability is a physical probability. It is at least a matter of linguistic convenience or consistency, and it may be more.

A full discussion of the relationships between the various kinds of probability would take us too far afield. I shall merely repeat dogmatically my opinion that although there are at least five different kinds of probability we can get along with just one kind, namely, subjective probability. This opinion is analogous to the one that we can know the world only through our own sensations, an opinion that does not necessarily make us solipsists, nor does it prevent us from talking about the outside world. Likewise, the subjectivist can be quite happy talking about physical probability, although he can measure it only with the help of subjective probability.

Bearing on Indeterminism

On the face of it, the assumption that physical probabilities exist seems to imply the metaphysical theory of indeterminism. I shall conclude by trying to analyze this opinion.

When I say that a theory is "metaphysical," I mean that there is no conceivable experiment that can greatly change the logarithm of its odds. (The odds corresponding to probability p are defined as $p/(1-p)$. It lies between 0 and plus infinity, and its logarithm lies between $-\infty$ and $+\infty$.) No theory is metaphysical if it can be virtually either proved or falsified, because its log-odds would then become very large, positive or negative. According to this definition, it is a question of degree whether a theory is metaphysical.

For example, the theory of determinism is less credible than it was a hundred years ago, but is by no means disproved and never will be. A statistician can never prove that "random numbers" are not "pseudo-random," and likewise "pseudo-indeterminism" cannot be disproved (*10*).

We can consistently talk about physical probability without committing ourselves to the metaphysical theory that the universe is indeterministic, but only if we accept the existence of subjective probability or credibility. For if we assume determinism we can get physical probabilities only by having an incompletely specified physical setup. In this incomplete specification there must be probabilities. If we are determinists we must attribute these latter probabilities to our own ignorance and not merely to something basic in nature "out there." Whether or not we assume determinism, every physical probability *can* be interpreted as a subjective probability or as a credibility. If we do assume determinism, then such an interpretation is forced upon us.

Those philosophers who believe that the only kind of probability is physical must be indeterminists. It was for this reason that von Mises asserted indeterminism before it became fashionable. He was lucky.

References and Notes

1. J. M. Keynes, *A Treatise on Probability* (Macmillan, London and New York, 1921; St. Martin's Press, New York, 1952).
2. H. Jeffreys, *Theory of Probability* (Oxford Univ. Press, New York and London, 1939; ed. 2, 1948).
3. F. P. Ramsey, *Foundations of Mathematics* (Harcourt, London, 1931; Humanities Press, New York, 1950).
4. B. de Finetti, *Ann. inst. Henri Poincaré* **7**, 1 (1937).
5. B. O. Koopman, *Ann. Math. Ser. 2*, **41**, 269 (1940); *Bull. Am. Math. Soc.* **46**, 763 (1940).
6. R. Carnap, *Logical Foundations of Probability* (Univ. of Chicago Press, Chicago, 1950).
7. B. Russell, *Human Knowledge, Its Scope and Limits* (Simon and Schuster, New York, 1948), part V.
8. I. J. Good, *Probability and the Weighing of Evidence* (Hafner, London and New York, 195→ *J. Roy. Statist. Soc. Ser. B* **14**, 107 (1952); in *Uncertainty and Business Decisions*, G. L. S. Shackle, Ed. (Univ. Press of Liverpool, Liverpool, 1957), ed. 2, chap. 3; *J. Inst. Actuaries* 82, 249 (1956); "Significance tests in series and in parallel," *J. Am. Statist. Assoc.*, in press.
9. L. J. Savage, *Foundations of Statistics* (Wiley, New York and London, 1954).
10. I. J. Good, *J. Roy. Statist. Soc. Ser. B* **20**, 35 (1958).

Water Transport

This classical problem in plant physiology is becoming increasingly amenable to mathematical analysis.

James Bonner

Plant physiology, even though it has existed as an organized science for one hundred years, still has its classical problems—problems which have been studied by many investigators during this hundred years, and problems which are nonetheless still unsolved. Such a classical problem of plant physiology is that of water transport. It is not, in fact, unsolved in principle today. Certainly the question of how water ascends the trunk of the tree to supply the transpiring leaves has been solved in principle by Dixon (*1*) and by Renner (*2*). Although the tension-cohesion hypothesis of water transport proposed by Dixon (*1*) has been attacked from time to time, it has, I believe, thus far always turned out that the attackers have been barking up the wrong tree. In a broader sense, however, "water transport" can be used to mean material transport of water to, within, and from the plant, and in this sense water transport bristles with unsolved and even with unposed questions.

In this article I propose to take up the successive steps in the material transport of water and to comment for each step on recent contributions which appear to be of importance, as well as upon problems which appear to pose further interesting questions.

From Soil to Root

Let us first consider water movement from soil to root. As in all cases of water movement, this consists of water flow from regions of lower diffusion pressure deficit (DPD) to regions of higher DPD. The soil DPD is determined by soil moisture stress and by the content of osmotically active solutes in the soil water. But the solutes are, in general, salts which can be taken up by the root and increase the DPD of the root cells. It might therefore appear that such solutes would contribute little to the soil-plant DPD-gradient which determines water uptake by the root. Indeed Walter (*3*) many years ago declared explicitly that soil solutes which can permeate the root do not in fact play any role in moisture uptake by the plant.

Wadleigh (*4*) and his colleagues at Riverside (California) have, however, developed the concept of total soil moisture stress, a total made up of physically and osmotically determined components. These workers have shown experimentally that the addition of salts to soil does in fact increase the effective soil DPD against which the plant must work. We have therefore a paradox—Walter's view based on sound plant physiological foundations and Wadleigh's view based on experiment.

The paradox has been resolved by John Philip (*5*) of the agricultural physics group of the Commonwealth Scientific and Industrial Research Organization's Division of Plant Industry (formerly at Deniliquin, now at Canberra). By applying recent advances in the quantitative theory of water movement in soils, Philip has shown that, during even moderate transpiration, removal of water

The author is a member of the staff of the Division of Biology, California Institute of Technology, Pasadena. This article is adapted from an address made 26 Aug. 1958 at the annual meeting of the American Society of Plant Physiologists, held in Bloomington, Ind.