Probability

Alan Hájek

In *New Dictionary of the History of Ideas*, Charles Scribner's Sons, 2004.

"Probability is the very guide of life", just as Bishop Butler wrote in 1736. Probability judgments of the efficacy and side-effects of a pharmaceutical drug determine whether or not it is approved for release to the public. The fate of a defendant on trial for murder hinges on the jurors' opinions about the probabilistic weight of evidence. Geologists calculate the probability that an earthquake of a certain intensity will hit a given city, and engineers accordingly build skyscrapers with specified probabilities of withstanding such earthquakes. Probability undergirds even measurement itself, since the error bounds that accompany measurements are essentially probabilistic confidence intervals. We find probability wherever we find uncertainty—that is, almost everywhere in our lives.

It is surprising, then, that probability arrived comparatively late on the intellectual scene. To be sure, a notion of randomness was known to the ancients—Epicurus, and later Lucretius, believed that atoms occasionally underwent indeterministic swerves. Averroes had a notion of 'equipotency' that might be regarded as a precursor to probabilistic notions. But probability theory was not conceived until the 17th century, with the Fermat-Pascal correspondence and the *Port-Royal Logic*. Over the next three centuries, the theory was developed by such authors as Huygens, Bernoulli, Bayes, Laplace, Condorcet, de Moivre, Venn, Johnson, and Keynes. Arguably, the crowning achievement was Kolmogorov's axiomatization in 1933, which put probability on rigorous mathematical footing.

The formal theory of probability

In Kolmogorov's theory, probabilities are numerical values that are assigned to 'events'. The numbers are non-negative; they have a maximum value of 1; and the probability that one of two mutually exclusive events occurs is the sum of their individual probabilities. Said more formally: given a set Ω and a privileged set of subsets F of Ω, probability is a function *P* from F to the real numbers that obeys, for all *X* and Y in F:

[Editor: Put in a BOX entitled 'Kolmogorov's probability axioms'?:]

A1.	$P(X) \geq 0$.	(Non-negativity)
A2.	$P(\Omega) = 1$.	(Normalization)
A3.	$P(X \cup Y) = P(X) + P(Y)$ if $X \cap Y = \varnothing$.	(Additivity)

Kolmogovov goes on to give an infinite generalization of A3, so-called 'countable additivity'. He also defines the *conditional probability* of *A* given *B* by the formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (P(B) > 0).$$

Thus, we may say that the probability that the toss of a fair die results in a 6 is 1/6, but the probability that it results in a 6 *given* that it results in an even number, is $\frac{1/6}{1/2} = 1/3$.

Important consequences of these axioms include various forms of Bayes' theorem, notably:

$$P(H|E) = [P(H)/P(E)]\, P(E|H)$$

$$= P(H)P(E|H) / [P(H)P(E|H) + P(\sim H)P(E|(\sim H)]$$

This theorem provides the basis for Bayesian confirmation theory, which appeals to such probabilities in its account of the evidential support that a piece of evidence E provides a hypothesis H. $P(E|H)$ is called the 'likelihood' (the probability that the hypothesis gives to the evidence) and $P(H)$ the 'prior' probability of H (the probability of the hypothesis in the absence of any evidence whatsoever).

Events A and B are said to be *independent* if $P(A \cap B) = P(A)P(B)$. If $P(A)$ and $P(B) > 0$, this is equivalent to $P(A|B) = P(A)$ and to $P(B|A) = P(B)$—intuitively, information about the occurrence of one of the events does not alter the probability of the other. Thus, the outcome of a particular coin toss is presumably independent of the result of the next presidential election. Independence plays a central role in probability theory. For example, it underpins the various important laws of large numbers, whose content is roughly that certain well-behaved processes are very likely in the long run to yield frequencies that would be 'expected' on the basis of their probabilities.

While the mathematics of Kolmogorov's probability theory is well understood and thoroughly developed (Feller 1968 is a classic text), its *interpretation* remains controversial. We thus turn to several rival accounts of what probabilities are, and how they are to be determined (see Hájek 2002 for more detailed discussion).

Interpretations of probability

The *classical* interpretation, historically the first, can be found in the works of Pascal, Huygens, Bernoulli, and Leibniz, and it was famously presented by Laplace (1814). It assigns probabilities in the absence of any evidence, or in the presence of symmetrically balanced evidence. In such circumstances, probability is shared equally among all the possible outcomes—the so-called ‘principle of indifference’. Thus, the classical probability of an event is simply the fraction of the total number of possibilities in which the event occurs. This interpretation was inspired by, and typically applied to, games of chance that by their very design create such circumstances—for example, the classical probability of a fair die landing with an even number showing up is 3/6.

Notoriously, the interpretation falters when there are competing sets of possible outcomes. What is the probability that the die lands 6 when tossed? If we list the possible outcomes as {1, 2, 3, 4, 5, 6}, the answer appears to be 1/6; but if we list them as {6, not-6}, the answer appears to be 1/2.

The *logical* interpretation retains the classical interpretation's idea that probabilities are determined *a priori* by the space of possibilities. But the logical interpretation is more general in two important ways: the possibilities may be assigned *unequal* weights, and probabilities can be computed *whatever* the evidence may be, symmetrically balanced or not. Indeed, the logical interpretation seeks to determine universally the degree of support or confirmation that a piece of evidence *E* confers upon a given hypothesis *H*. Carnap (1950) thus hoped to offer an ‘inductive logic’ that generalized deductive logic and its relation of ‘implication’ (the strongest relation of support).

A central problem with Carnap's program is that changing the language in which items of evidence and hypotheses are expressed will typically change the confirmation relations between them. Moreover, deductive logic can be characterized purely syntactically: it can be determined whether *E* implies *H*, or whether *H* is a tautology, merely by inspection of their symbolic structure, irrespective of their content. But Goodman showed that inductive logic must be sensitive to the meanings of words, for syntactically parallel inferences can differ wildly in their inductive strength. So inductive logic is apparently not of a piece with deductive logic after all.

Frequency interpretations date back to Venn (1876). Gamblers, actuaries and scientists have long understood that relative frequencies bear an intimate relationship to probabilities. Frequency interpretations posit the most intimate relationship of all: identity. Thus, the probability of 'heads' on a coin that lands heads 7 times out of 10 tosses is 7/10. In general:

the probability of an outcome *A* in a reference class *B* is the proportion of occurrences of *A* within *B*.

Frequentism still has the ascendancy among scientists who seek to capture an objective notion of probability, heedless of anyone's beliefs. It is also the philosophical position that lies in the background of the classical Fisher/Neyman-Pearson approach that is used in most statistics textbooks. Frequentism does, however, face some major objections. For example, a coin that is tossed exactly once yields a relative frequency of heads of either 0 or 1, whatever its true bias—the infamous 'problem of the single case'. Some frequentists (notably Reichenbach, and von Mises 1957) go on to consider infinite reference classes of hypothetical occurrences. Probabilities are then defined as limiting relative frequencies in suitable infinite sequences of trials. If there are in fact only finitely many trials of the relevant type, then this requires the actual sequence to be extended to a hypothetical or 'virtual' sequence. This creates new difficulties. For instance, there is apparently no fact of the matter of how the coin in my pocket would have landed if it had been tossed once, let alone indefinitely. A well-known problem for any version of frequentism is that relative frequencies must be relativized to a reference class. Suppose that you are interested in the probability that you will live to age 80. Which reference class should you consult? The class of all people? All people of your gender? All people who share your lifestyle? ... Only you have all these properties; but then the problem of the single case returns.

Propensity interpretations, like frequency interpretations, regard probability as an objective feature of the world. Probability is thought of as a physical propensity, or disposition, or tendency of a given type of physical situation to yield an outcome of a certain kind, or to yield a long run (perhaps infinite) relative frequency of such an outcome. This view, which originated with Popper (1959), was motivated by the desire to make sense of single-case probability attributions on which frequentism apparently foundered, particularly those found in quantum mechanics. See Gillies (2000) for a useful survey.

A prevalent objection is that it is not informative to be told that probabilities are 'propensities'. For example, what exactly is the property in virtue of which this coin has a 'propensity' of 1/2 of landing heads (when suitably tossed)? Indeed, some authors regard it as mysterious whether propensities even obey the axioms of probability in the first place. To the extent that propensity theories are parasitic on long run frequencies, they also seem to inherit some of the problems of frequentism.

Subjectivist interpretations, pioneered by Ramsey (1926) and de Finetti (1937), regard probabilities as *degrees of belief,* or *credences* of appropriate agents. These agents cannot be actual people since, as psychologists have repeatedly shown, people typically violate probability theory in various ways, often spectacularly so. Instead, we imagine the agents to be ideally rational. Ramsey thus regarded probability theory to be the 'logic of partial belief'. Underpinning subjectivism are so-called 'Dutch Book arguments'. They begin by identifying agents' degrees of belief with their betting dispositions, and they then show that anyone whose degrees of belief violate the axioms of probability is 'incoherent'—susceptible to guaranteed losses at the hands of a cunning bettor. Equally important, but often neglected, is the converse theorem that obedience to the probability axioms protects one from such an ill fate. Subjectivism has proven to be influential especially among social scientists, 'Bayesian' statisticians, and philosophers.

A more general approach, again originating with Ramsey, begins with certain axioms on rational preferences—for example, if you prefer A to B and prefer B to C, then you prefer A to C. It can be shown that if you obey these axioms, then you can be 'represented' by a probability function (encapsulating your credences about various propositions) and a utility function (encapsulating your strengths of desire for these propositions). This means that you will rate the choice-worthiness of an option open to you according to its 'expected utility', a weighted average of the various possible utilities associated with that action, the corresponding probabilities providing the weights. This is the centerpiece of decision theory.

Radical subjectivists such as de Finetti recognize no constraints on initial, or 'prior' subjective probabilities beyond their conformity to axioms A1–A3. But they typically advocate a learning rule for updating probabilities in the light of new evidence. Suppose that initially you have credences given by a probability function $P_{initial}$, and that you become certain of E (where E is the strongest such proposition). What should be your new probability function P_{new}? The favored updating rule among Bayesians is conditionalization; P_{new} is related to $P_{initial}$ as follows:

(Conditionalization) $P_{new}(X) = P_{initial}(X|E)$ (provided $P_{initial}(E) > 0$).

Radical subjectivism has faced the charge of being too permissive. It apparently licenses credences that we would ordinarily regard as crazy. For example, you can assign without its censure probability 0.999 to your being the only thinking thing in the universe—provided that you remain coherent (and update by conditionalization). It also seems to allow inference rules that are considered fallacious, such as the gambler's fallacy (believing, for instance, that after a surprisingly long run of heads, a fair coin is more likely to land tails). A standard defence (e.g., Howson and Urbach 1993) appeals to famous 'convergence-to-truth', and 'merger-of-opinion' results. Their upshot is that in the long run, the effect of choosing one prior rather than another is attenuated: successive conditionalizations on the evidence will, with probability one, make a given agent eventually converge to the truth, and thus initially discrepant agents eventually come to agreement. Some authors object that these theorems tell us nothing about how quickly the convergence occurs; in particular, they do not explain the unanimity that we in fact often reach, and often rather rapidly.

Some recent developments

In response, certain subjectivists nowadays are more demanding, adding further constraints to their subjectivism. For example, we might evaluate credences according to how closely they match the corresponding relative frequencies: how well 'calibrated' they are. 'Scoring rules' that refine calibration have also been canvassed. Various subjectivists believe that rational credences are guided by objective chances (perhaps thought of as propensities), so that if a rational agent knows the objective chance of a given outcome, her degree of belief will be the same. There has been important research on the aggregation of opinions and preferences of multiple agents. This problem is well known to aficionados of the risk-assessment literature. Moreover, in light of work in the economics and psychology literature on 'bounded rationality', there have been various attempts to 'humanize' Bayesianism—for example, in the study of 'degrees of incoherence', and of vague probability and decision theory (in which credences need not assume precise values).

Recent times have also seen attempts to rehabilitate the classical and logical interpretations, and in particular the principle of indifference. Some 'objective' Bayesians appeal to information theory, arguing that prior probabilities should maximize 'entropy' (a measure of the 'flatness' of a probability distribution), subject to the constraints of a given problem. Probability theory has also been influenced by advances in theories of randomness and complexity theory (see Fine 1973, Li and Vitanyi 1997), and approaches to the 'curve-fitting' problem—familiar in the computer science, artificial intelligence, and philosophy of science literature—that attempt to measure the simplicity of theories.

While Kolmogorov's theory remains the orthodoxy, a host of alternative theories of probability have been developed (see Fine 1973, Mückenheim et al. 1986). For instance, there has been increased interest in non-additive theories, and the status of countable additivity is the subject of lively debate. Some authors have proposed theories of primitive conditional probability functions, in which conditional probability replaces unconditional probability as the fundamental concept. Fertile connections between probability and logic have been explored under the rubric of 'probabilistic semantics' or 'probability logic'.

Some applications of probability

Probability theory thus continues to be a vigorous area of research. Moreover, its advances have myriad ramifications. Probability is explicitly used in many of our best scientific theories—for example, quantum mechanics and statistical mechanics. It is also implicit in much of our theorizing. A central notion in evolutionary biology is 'fitness', or expected number of offspring. Psychologists publish their conclusions with significance levels attached. Agricultural scientists perform analyses of variance on the efficacy of fertilizers on crop yields. Economists model currency exchange rates over time as stochastic processes: sequences of random variables. In cognitive science and philosophy, probability functions model states of opinion. Since probability theory is at the heart of decision theory, it has consequences for ethics and political philosophy. And assuming, as many authors do, that decision theory provides a good model of rational decision-making, it apparently has implications for even mundane aspects of our daily lives.

In short, probability is virtually ubiquitous. Bishop Butler's dictum is truer today than ever.

Bibliography

Butler, Joseph (1961): *Analogy of Religion*, 1736; reprinted by Frederick Ungar Publishing Company.

Carnap, Rudolf (1950): *Logical Foundations of Probability*, Chicago: University of Chicago Press.

De Finetti, Bruno (1937): “La Prévision: Ses Lois Logiques, Ses Sources Subjectives.” *Annales de l’Institut Henri Poincaré*, 7, 1-68; translated as “Foresight. Its Logical Laws, Its Subjective Sources”, in *Studies in Subjective Probability*, eds. H. E. Kyburg, Jr. and H. E. Smokler. New York: Robert E. Krieger Publishing Co., 1980.

Feller, William (1968): *An Introduction to Probability Theory and Its Applications*, 3rd ed., John Wiley & Sons, Inc..

Fine, Terrence (1973): *Theories of Probability*, New York: Academic Press.

Gillies, Donald (2000) “Varieties of Propensity*”, British Journal for the Philosophy of Science,* 51, 807-835.

Hájek, Alan (2002): “Probability, Interpretations of”, *The Stanford Encyclopedia of Philosophy,* ed. E. Zalta, *http://plato.stanford.edu/entries/probability-interpret/.*

Howson, Colin and Peter Urbach (1993): *Scientific Reasoning: The Bayesian Approach*, 2nd ed., Open Court, Chicago.

Kolmogorov, Andrei. N. (1933): *Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse Der Mathematik;* translated as *Foundations of Probability*, New York: Chelsea Publishing Company, 1950.

Laplace, Pierre Simon (1814): *A Philosophical Essay on Probabilities,* New York: Dover Publications Inc., English edition 1951.

Li, Ming and Paul Vitanyi (1997): *An Introduction to Kolmogorov Complexity and its Applications*, 2nd ed., New York: Springer-Verlag.

Muckenheim, W., Ludwig, G., Dewdney, C., Holland, P., Kyprianidis A., Vigier, J., Petroni, N., Bartlett, M., and Jaynes. E. (1986): "A Review of Extended Probability". *Physics Reports*, 133, 337–401.

Popper, Karl (1959): *The Logic of Scientific Discovery*, London: Hutchinson & Co..

Ramsey, Frank P., (1926): "Truth and Probability." In *Foundations of Mathematics and other Essays*, ed. R. B. Braithwaite, 156-198, Routledge & P. Kegan, 1931; reprinted in *Studies in Subjective Probability*, eds. H. E. Kyburg, Jr. and H. E. Smokler, 2nd ed., 23-52, New York: R. E. Krieger Publishing Co., 1980; reprinted in *Philosophical Papers*, ed. D. H. Mellor. Cambridge: University Press, Cambridge, 1990.

Venn, John (1876): *The Logic of Chance*, 2nd ed. Macmillan and co; reprinted, New York, 1962.

von Mises, Richard (1957): *Probability, Statistics and Truth*, revised English edition, New York: Macmillan.