

CHAPTER 2

THE MEANING OF PROBABILITY

INTRODUCTION

by
Glenn Shafer

The meaning of probability has been debated since the mathematical theory of probability was formulated in the late 1600s. The five articles in this section have been selected to provide perspective on the history and present state of this debate.

Mathematical statistics provided the main arena for debating the meaning of probability during the nineteenth and early twentieth centuries. The debate was conducted mainly between two camps, the subjectivists and the frequentists. The subjectivists contended that the probability of an event is the degree to which someone believes it, as indicated by their willingness to bet or take other actions. The frequentists contended that probability of an event is the frequency with which it occurs. Leonard J. Savage (1917-1971), the author of our first article, was an influential subjectivist. Bradley Efron, the author of our second article, is a leading contemporary frequentist.

A newer debate, dating only from the 1950s and conducted more by psychologists and economists than by statisticians, has been concerned with whether the rules of probability are descriptive of human behavior or instead normative for human and machine reasoning. This debate has inspired empirical studies of the ways people violate the rules. In our third article, Amos Tversky and Daniel Kahneman report on some of the results of these studies.

In our fourth article, Amos Tversky and I propose that we resolve both debates by formalizing a constructive interpretation of probability. According to this interpretation, probabilities are degrees of belief deliberately constructed and adopted on the basis of evidence, and frequencies are only one among many types of evidence. The standard rules for probability, which are followed by frequencies, will be followed by degrees of belief based directly on frequencies, and by degrees of belief assessed using an analogy to the calculus of frequencies, but they may be violated by other degrees of belief. Thus these standard rules are neither descriptive (people do not usually obey them precisely) nor absolutely normative (it is sometimes reasonable not to obey them).

Our fifth article, by Judea Pearl, Dan Geiger and Thomas Verma, shows how conditional independence can be studied axiomatically. This topic has not traditionally been given prominence in the study of probability, but as we saw in Chapter 1, AI has brought it to the forefront.

The remainder of this introduction is concerned with the evolution of the frequentist/subjectivist debate and its relevance to AI, the current diversity of opinion about the meaning of probability in various intellectual communities, the descriptive/normative debate, and the constructive interpretation.

1. Frequentists vs. Subjectivists: A Historical Perspective

Students of probability have been divided into subjectivists and frequentists since the mid-nineteenth century, but the debate between the two camps has become relevant to practical applications only in the twentieth century.

The unchallenged nineteenth-century authority on the use of probability was Laplace, whose *Théorie analytique des probabilités* was first published in 1812. Laplace subscribed to a subjective definition of probability, and he used the method of inverse probability, which first appeared in the work of Thomas Bayes, posthumously published in 1763. But most of Laplace's statistical methods were acceptable to nineteenth-century empiricists who preferred a definition in terms of frequencies. The frequentists did criticize the subjectivism involved in the application of inverse probability to questions of human testimony, but this was a minor issue. The frequentists had no objections to the most successful statistical work of the time, the use of least-squares in astronomy. Least squares could be justified by frequentist arguments as well as by inverse probability (Stigler 1986).

In the twentieth century, however, new frequentist statistical techniques were developed that were subtle and powerful enough to be useful in biology and the social sciences, and these frequentist

techniques did not have obvious subjectivist interpretations. Consequently, frequentism became the dominant philosophy of probability among mathematical statisticians and scientists. The English statistician and geneticist R.A. Fisher (1890-1962) was particularly influential in this development. Fisher, whose work on maximum-likelihood estimation, significance testing, the analysis of variance, and the design of experiments still dominates much of applied statistics, was outspokenly critical of subjectivist ideas and of Bayes's theorem.

As the new statistical theory matured, and its own problems and paradoxes became clearer, the pendulum began to swing back to subjective ideas. This development is often dated from the publication of L.J. Savage's influential book, *The Foundations of Statistics*, in 1954. In this book, Savage gave a set of axioms for a person's preferences among acts. Rationality, he said, demands that a person's preferences obey these axioms. Savage proved that if preferences do obey his axioms, then these preferences are identical with preferences calculated from some set of numerical probabilities and utilities. In other words, there exist probabilities for possible states of the world and utilities for possible consequences of acts such that the preferences agree with the ranking of acts obtained by calculating their expected utility using these probabilities and utilities. Since there were no frequencies in the picture, the probabilities that emerged in this way had to be subjective. Savage concluded that a rational person should have subjective probabilities, and he went on to try to justify twentieth-century statistical methods on the basis of these subjective probabilities.

Savage's joint axiomatization of subjective probability and utility was inspired by the earlier axiomatization of utility published by John von Neumann and Oskar Morgenstern in 1947, in the second edition of *Theory of Games and Economic Behavior*. Von Neumann and Morgenstern assumed the existence of known objective probabilities, but they remarked in a footnote that a generalization to subjective probabilities should be possible.

Savage borrowed more than just a technical idea from economics. He also borrowed a viewpoint that allowed him to defend subjective probability within the empiricist philosophy that had inspired frequentism. The frequentists had scorned "degree of belief" because it had no empirical content. The economists, however, conceived of themselves as studying the behavior of people, a respectable empirical endeavor. Savage's postulates were postulates about behavior.

When he wrote his book, Savage was interested in subjective ideas as a new and better foundation for current statistical practice. But he and other subjectivists soon realized that many frequentist methods, such as minimax rules, tail-area tests, and tolerance intervals, could not be squared with a subjectivist viewpoint. They also realized that genuinely subjectivist alternatives were often theoretically possible but difficult to implement. Implementations would involve the development of complicated models, the assessment of a large number of "prior" probabilities, and a great deal of computation. Gradually, however, focus shifted from philosophical debate to serious work on such implementations; this was just beginning to happen when Savage wrote the article that is reprinted in this chapter. Today, the Bayesians, as the subjectivists have been called since the 1960s, are responsible for much of the theoretical work published in journals of mathematical statistics and are making inroads into statistical practice. Though most scientists and scientific journals still insist on conventional frequentist treatments of data, the frequentists now hold only an uneasy upper hand over their Bayesian colleagues within the community of mathematical statisticians.

One reason for the continued popularity of frequentist methods is that mathematical statistics tends to define itself in terms of the analysis of frequency data. This is what gives the field a role in the scientific world and an identity independent of pure mathematics. Most Bayesian statisticians compromise on the meaning of probability; they agree that their goal is to estimate objective probabilities from frequency data, but they advocate using subjective prior probabilities to improve the estimates (Good 1983).

In addition to the frequent and the subjective Bayesian views on the meaning of probability, there is a third established view, which attributes non-frequentist but objective interpretations to prior probabilities. Savage calls this "necessary" view; Efron calls it "objective Bayesian." It can be regarded as a continuation of the classical view of Laplace (Daston 1988), who saw probability as rational degree of belief, not as the degree of belief of an actual person. Those usually associated with this view include John Maynard Keynes (1921), and Harold Jeffreys (1939), and Richard T. Cox (1961; see also Cox's article in Chapter 6).

For further information about the debate over the meaning of probability, see Nagel (1939), Kyburg and Smokler (1964), Barnett (1973), and Fine (1973). For more information on technical aspects of the debate within mathematical statistics, see Efron (1978).

2. Is the Traditional Debate Relevant to AI?

Since artificial intelligence has made limited use of mathematical statistics, some readers might question whether the traditional debate between frequentists and subjectivists is relevant to AI. But clear-headed use of probability in any field, including AI, requires some understanding of the meaning of probability. Even though the issues in AI are not the same as the issues in mathematical statistics, an understanding of the debate in mathematical statistics can advance efforts to find probability semantics appropriate for AI. Moreover, an appreciation of the distinction between frequentist and subjectivist ideas is essential to assessing the potential of probability in AI. The frequentist claim that probability is appropriate when frequency data are available must be distinguished from the broader Bayesian claims for probability in AI.

One way in which the debate in mathematical statistics must be modified for application to AI is in its identification of subjectivity with prior probabilities. As I have already pointed out, Bayesian and frequentist statisticians tend to agree on the objective, frequentist character of the statistical model for the data that the statistician gathers. This model is usually only partially known, however, and the two schools have differed on how to handle this lack of knowledge. The Bayesians prefer to assess prior subjective probabilities for the different possible statistical models and then use the data to update these prior probabilities to posterior probabilities, while the frequentists prefer to rely on the data alone to estimate the model.

In many everyday problems of interest to AI, the kind of data analyzed by statisticians is not available, and hence the idea of an objective model is not so natural. In these problems, the frequentist has little to say, but the Bayesian is free to produce analyses in which both the model and the priors are subjective.

The stereotypical situation of mathematical statistics, where the prior probabilities are subjective but the model is objective, is often reversed in everyday problems. We may have frequencies on which to base prior probabilities, but no frequencies on which to base a model. Examples of this are provided by Tversky and Kahneman in their article in this chapter. One example involves a person drawn from a population of lawyers and engineers. People are told how many lawyers and engineers are in the population, they are given a description of a person randomly drawn from the population, and they are asked to assess probabilities that this person is a lawyer or an engineer. Here the known frequencies provide prior probabilities, but the model that specifies whether a lawyer or engineer is more likely to match the description is relatively subjective. When they answer Tversky and Kahneman's question, people tend to ignore the frequencies in the population and rely entirely on their subjective models of lawyers and engineers. Thus the stereotypical preference for objective rather than subjective information is also reversed.

Reliance on subjective models is natural for human beings, since we have had little means for storing frequency information in the course of our evolutionary experience. We have learned instead to store information in terms of causal models. There is considerable evidence, however, that our neglect of frequency information is often costly (Dawes, Faust, and Meehl 1989). Thus the fact that human beings tend not to use frequency information or statistical methods is not a persuasive argument against their use in artificial intelligence.

3. The Growing Diversity of Probability

While the twentieth century has seen the probability controversy evolve along one line in statistics, it has also seen probability grow in importance in a number of other fields. The different preoccupations of these fields have resulted in different attitudes towards frequentism and subjectivism. This section will briefly review the situation in decision analysis, applied probability, economics, and physics.

Decision Analysis. This field is treated in detail in the next chapter, but it deserves to be mentioned here because it is the result of interest in subjective expected utility spilling outside the realm of statistics. There are decision analysts in psychology, engineering, management science, medicine, and even law, all seeking to help people express their wants in terms of numerical utilities, express their beliefs in terms of subjective probabilities, use further evidence to sharpen these probabilities, and make decisions by calculating expected utilities. Though they are willing to use frequencies, they emphasize instead the measurement, elicitation or construction of personal opinion. Hence the probabilities they use are unequivocally subjective.

Applied Probability. In many areas of engineering, medicine, business, and the sciences, there is interest in objective probability models that are too complex to be estimated in detail from frequency data but are sufficiently realistic to give useful rough or qualitative predictions. The enterprise of constructing such models is often called “applied probability” (Ross 1970).

Applied probability uses a variety of probability models. Most of them involve change over time, and when this is emphasized, the models are called “stochastic processes.” When there is also an emphasis on the interaction of a small number of related processes, the models are sometimes called “queuing models”; such models have been applied to the management of dams, transportation systems, and inventories, to the design of manufacturing processes and computer systems, and to many other problems. When there is an emphasis on spatial relations, as in many problems in the natural sciences, other classes of models enter, including diffusion processes and point processes.

In cases where substantial frequency data can be obtained, so that the probabilities in the models can be estimated accurately, applied probability begins to look like part of mathematical statistics; stochastic processes become “time series.” In many cases, however, there is insufficient data for precise estimation, either because repetition of the physical processes is impossible, or because the models are too complex for estimation to be precise with any reasonable amount of data.

Because of their emphasis on physical processes, as opposed to the processes of human choice that interest decision analysts and economists, applied probabilists usually have an objective, frequentist understanding of their probabilities.

Economics. Theoretical economics has been one of the most important areas of growth for probability in the past several decades. Increasingly, economists have used probability not only to analyze data but also to model people's behavior.

Before von Neumann and Morgenstern's axiomatization of utility, most microeconomic models were based on the assumption of perfect information. The legitimacy that von Neumann and Morgenstern gave to expected utility encouraged mathematical economists to relax this assumption and allow for uncertainty, and during and since the 1960s there has been an avalanche of work on microeconomic models that use probability (Diamond and Rothschild 1978). In more recent years, such models have also become important in macroeconomics (Lucas and Sargent 1981, von Furstenberg and Jeong 1988).

In most of this work, the probabilities are explicitly subjective. They are the probabilities of the economic agents in the economists' models, probabilities that these agents use to make choices. In some cases they are also frequencies or objective probabilities; they may be frequencies that the agents happen to know and decide to adopt as their degrees of belief. But this is secondary. What is essential is that they represent the beliefs of the people being modelled.

Objective probabilities do appear in some economic models. They play a natural role in modelling markets where frequencies are salient, such as the market for insurance. But this has become secondary to the use of probabilities to model people's internal decision-making.

Physics. Probability already played an important role in physics in the last century, in statistical mechanics. In this century it has come to play an apparently more fundamental role, in quantum mechanics.

Superficially, at least, all the probabilities in physics have a clear frequentist interpretation. Physicists verify that they are correct by repetitive experiments. Some philosophically-minded physicists have offered subjectivist interpretations, however, especially in the case of quantum mechanics, where Heisenberg's uncertainty principle clouds the objectivity of the theory (von Neumann 1955). Partly because of confusion about the foundations of quantum mechanics, and partly because probability in quantum mechanics has a distinctive mathematical form not found outside physics, the example of quantum mechanics has had relatively little influence on the debate between frequentism and subjectivism in other fields.

Interestingly, physicists who turn to using probability in other fields have been more likely than others scholars to adopt the necessary or objective Bayesian view. Harold Jeffreys (1939), E.T. Jaynes (1968), and Richard T. Cox (1961) were all physicists. Probabilists in AI who were originally trained in physics have also shown an affinity for the objective Bayesian view.

This affinity may be due to the fact that probabilities are often derived from physical theory. Experiments serve to test the theory and possibly to correct the probabilities, as when it is found that photons obey Bose-Einstein statistics rather than Fermi-Dirac statistics (Feller 1968). Physicists who turn their attention to other fields are tempted to think that it should be possible to derive *a priori* the prior subjective probabilities needed for Bayesian statistical inference, even if there are no

experiments that could correct or verify these probabilities. They therefore defend “objective priors” based on symmetries or entropy calculations.

4. *Is Probability Theory Normative or Descriptive?*

The distinction between “normative” and “descriptive” interpretations of probability theory derives from the interplay between the economic concerns of von Neumann and Morgenstern and the statistical concerns of Savage.

Economic theories are about the behavior of people. These models are usually built on the assumption that people act in their own economic self interest, and though economists acknowledge that people sometimes violate this assumption, they seek out for their work situations where people follow the assumption closely enough that theories built on it can accurately predict economic behavior. It is at this level, the level of predictions about aggregate behavior, that economic theories are usually tested. More direct tests of whether individuals act in their own self interest are considered less interesting by economists.

Von Neumann and Morgenstern presented their utility theory in the usual spirit of economics. It differed from most of the economic theory of its time, however, in that it undertook to explain individual behavior. Most microeconomic theory of the time was content to begin with assumptions such as the law of demand: there will be more buyers if the price is lower. Utility theory attempted to push economic explanation deeper, by explaining why an individual will buy more when the price is lower.

When Savage formulated his own axioms for subjective probability and utility, he hoped that these axioms too would be sufficiently true of people to be of interest to economic theory. His main interest was in statistical theory, however, and he was uneasy about using an unestablished empirical hypothesis as a foundation for statistics. So he borrowed from the philosophy of logic a distinction that had not appeared before in probability, the distinction between “normative” and “descriptive” interpretations. He declared that his axioms defined rational behavior. Whether his axioms were descriptive - whether people really were rational - was an interesting empirical question. But statisticians and other people making decisions should try to be rational in any case, so it was normative for statisticians to use subjective expected utility.

The empirical question about individual behavior that Savage formulated was also raised by other economists in the early 1950s. Did people really obey von Neumann and Morgenstern's and Savage's axioms? This question soon attracted the interest of experimental psychologists, a group not usually interested in economic theory. For these scholars, the probability and utility axioms provided a unique opportunity to subject economic theory to direct empirical test.

The psychologists who undertook to test the axioms were sympathetic with utility theory; like scholars in many fields, they were charmed by its mathematical elegance and power. Their early work, in the 1950s, tended to confirm the rough accuracy of the axioms. But as the issue was studied in more detail and with more ingenuity, attention focused on ways in which people violate the axioms.

The psychologists have now been at work on this issue for over two decades. The articles by Tversky and Kahneman in this and the next chapter provide a mature overview of the results (see also Kahneman, Slovic, and Tversky 1982). One way of summarizing these results is to say that people frequently violate the axioms. Another way is to say that they often do not have the well-defined preferences that the axioms are about. They will humor psychologists who ask them to make choices on questionnaires, but these choices are often whimsical.

These negative results have been received very differently by the different communities interested in subjective expected utility. Bayesian statisticians and decision analysts have taken the escape route already laid out by Savage. Yes, they have said, the axioms are descriptively wrong; people do not behave as they should. But the axioms are nonetheless normative.

The reaction of economists has been more complicated and diverse. Many, following Friedman (1953), have argued that economic models should be tested only by their predictions about aggregate behavior; direct tests of assumptions about the rationality of individuals are pointless, since it is generally impossible to gauge how important deviations from rationality are to aggregate predictions. A few, following Simon (1986), have argued that the looseness of the link between individual rationality and aggregate behavior means that the success of prediction cannot vindicate the assumptions of individual rationality; emphasis should be placed instead on the parts of the model closer to the predictions. Others, such as Plott (1986), have argued for experimental work complex enough to test both deviations from rationality and their importance.

5. *Constructive Probability*

Applied statisticians and decision theorists who do not accept the normative claims made for Bayesian methods sometimes use these methods, but they do so cautiously. If they do not feel that subjective probabilities are sufficiently supported by the evidence they are using, then they may use non-Bayesian methods instead.

This constructive, non-normative attitude can be made into a formal interpretation of probability, on a par with the frequentist and subjective interpretations. According to this constructive interpretation, the theory of probability describes an ideal situation, a situation like a game of chance in which the chances are known to the players. In this ideal situation, the probability of an event is simultaneously the frequency of the event, the fair price for a gamble on the event, and a measure of the degree to which one should believe that the event will happen. Application of probability means relating this ideal picture to a practical problem in some way, and the ways are many and diverse. Sometimes we simply use the ideal situation as a standard of comparison, as when we compare the predictions of weather or financial forecasters to the success that could be achieved by guessing at random. Sometimes we deliberately intertwine the ideal situation in a practical problem, as when we use random numbers to assign experimental treatments or to select individuals for a sample. Sometimes, as in Bayesian arguments, we draw an analogy between a practical problem and the ideal situation. We compare the practical problem with a game that uses dice with known frequency properties. We make individual probability judgments by drawing analogies between individual items of evidence in the two situations, and we combine probability judgments by drawing an analogy between the structure of the evidence in the two situations.

If we follow this constructive interpretation, then we must regard all probability arguments, Bayesian or non-Bayesian, as subject to review and evaluation. In the case of a Bayesian argument, the strength and quality of each analogy must be evaluated.

The constructive interpretation of probability also opens the way to alternative subjective theories of probability, which differ from the Bayesian theory by using different analogies. The article by Shafer and Tversky in this section applies the constructive interpretation to both the Bayesian theory and the Dempster-Shafer theory.

The Dempster-Shafer theory is discussed further in Chapter 7 of these readings. The philosophy of constructive probability is elaborated in Shafer (1981, 1990).

References

- Barnett, Vic (1973). *Comparative Statistical Inference*. Wiley. Second edition, 1982.
- Bayes, Thomas (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53** 370-418. Reprinted with commentary by G. A. Barnard in *Biometrika* **45** 293-315.
- Cox, Richard T. (1961). *The Algebra of Probable Inference*. The Johns Hopkins Press.
- Daston, Lorraine (1988). *Classical Probability in the Enlightenment*. Princeton University Press.
- Dawes, Robyn M., David Faust, and Paul E. Meehl (1989). Clinical versus actuarial judgment. *Science* **243** 1668-1674.
- Diamond, Peter, and Michael Rothschild (1978). *Uncertainty in Economics*. Academic Press.
- Efron, Bradley (1978). Controversies in the foundations of statistics. *American Mathematical Monthly* **85** 231-246.
- Feller, William (1968). *An Introduction to Probability Theory and Its Applications*. Volume I. Third edition. Wiley.
- Fine, Terrence L. (1973). *Theories of Probability; An Examination of Foundations*. Academic Press.
- Friedman, Milton (1953). *Essays in Positive Economics*. University of Chicago Press.
- Good, I.J. (1983). *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press.
- Jaynes, E.T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics* **4** 227-241.
- Jeffreys, Harold (1939). *Theory of Probability*. Oxford University Press. Second edition, 1948; third edition, 1961.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Keynes, John Maynard (1921). *A Treatise on Probability*. Macmillan: London.

- Kyburg, Henry E., Jr., and Howard E. Smokler, eds. (1964). *Studies in Subjective Probability*. Wiley. Second edition, 1980, Robert E. Krieger.
- Laplace (Pierre Simon, le Marquis de Laplace) (1812) . *Théorie Analytique des Probabilités*. Paris. Second edition, 1814; third edition, 1820.
- Lucas, Robert E., Jr., and Thomas J. Sargent, eds. (1981). *Rational Expectations and Econometric Practice*. Two volumes. University of Minnesota Press.
- Nagel, Ernest (1939). *Principles of the Theory of Probability*. (Volume 1, Number 6 of the *International Encyclopedia of Unified Science*) University of Chicago Press.
- Plott, Charles R. (1986). Rational choice in experimental markets. *Journal of Business* **59** S301-S327.
- Ross, Sheldon M. (1970). *Applied Probability Models with Optimization Applications*. Holden-Day.
- Shafer, Glenn (1981). Constructive probability. *Synthese* **48** 1-60.
- Shafer, Glenn (1990). The unity of probability. In *Acting Under Uncertainty: Multidisciplinary Conceptions*, George von Furstenberg, ed., Kluwer.
- Savage, Leonard J. (1954). *The Foundations of Statistics*. Wiley. Second revised edition published in 1972 by Dover.
- Simon, Herbert A. (1986). Rationality in psychology and economics. *Journal of Business* **59** S209-S224.
- Stigler, Stephen M. (1986). *The History of Statistics*. Harvard University Press.
- von Furstenberg, George M., and Jin-Ho Jeong (1988). Owning up to uncertainty in macroeconomics. *The Geneva Papers on Risk and Insurance*, Vol. 13, No. 46, pp. 12-90.
- von Neumann, John (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton University Press.
- von Neumann, John, and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press. Second edition, 1947; third edition, 1953.

Articles for Chapter 2

1. The Foundations of Statistics Reconsidered, by Leonard J. Savage. (1961) Originally published in 1961 on pp. 575-586 of *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, edited by Jerzy Neyman, University of California Press. Reprinted with some changes on pp. 173-188 of *Studies in Subjective Probability*, edited by Henry J. Kyburg, Jr., and Howard E. Smokler, Wiley, 1964. (The changes are minor; I noticed a new footnote, a correction in a reference number, and editing of the reference to another article that appeared in the symposium. The article does not appear in the second edition of Kyburg and Smokler's book, which was published by Krieger in 1980.)
2. Why Isn't Everyone a Bayesian? By Bradley Efron. (1986) *The American Statistician*, Vol. 40, No. 1, February 1986, pp. 1-5.
3. Judgment under Uncertainty: Heuristics and Biases, by Amos Tversky and Daniel Kahneman. (1974) *Science*, 27 September, Vol. 185, pp. 1124-1131.
4. Languages and Designs for Probability Judgment, by Glenn Shafer and Amos Tversky. (1985) *Cognitive Science*, Vol. 9, pp. 309-339.
5. Conditional Independence and Its Representations, by Judea Pearl, Dan Geiger, and Thomas Verma. (1989) *Kybernetika*, Vol. 25, No. 2, pp. 33-44.