Statistical Science
1990, Vol. 5, No. 4, 435–462

The Unity and Diversity of Probability

Glenn Shafer

1. INTRODUCTION

Mathematical probability and its child, mathematical statistics, are relative newcomers on the intellectual scene. Mathematical probability was invented in 1654 by two Frenchman, Blaise Pascal and Pierre Fermat. Mathematical statistics emerged from the work of the continental mathematicians Gauss and Laplace in the early 1800s, and it became widely useful only in this century, as the result of the work of three Englishmen, Francis Galton, Karl Pearson and R. A. Fisher.

In spite of these late beginnings, probability and statistics have acquired a dazzling range of applications. Inside the university, we see them taught and used in a remarkable range of disciplines. Statistics is used routinely in engineering, business, medicine and every social and natural science. It is making inroads in law and in the humanities. Probability, aside from its use in statistical theory, is finding new applications in engineering, computer science, economics, psychology and philosophy.

Outside the university, we see probability and statistics in use in a myriad of practical tasks. Physicians rely on computer programs that use probabilistic methods to interpret the results of some medical tests. The worker at the ready-mix company used a chart based on probability theory when he mixed the concrete for the foundation of my house, and the tax assessor used a statistical package on his personal computer to decide how much the house is worth.

In this article, I will sketch the intellectual history of the growth and diversification of probability theory. I will begin at the beginning, with the letters between the Parisian polymath Blaise Pascal and the Toulouse lawyer Pierre Fermat in 1654. I will explain how these authors, together with James Bernoulli, Abraham De Moivre and Pierre Simon, the Marquis de Laplace, invented a theory that unified the ideas of belief and frequency. I will explain how this unity crumbled under the assault of the empiricist philosophy of the nineteenth century, how the frequency interpretation of probability emerged from this assault and how a subjective (degree of belief) interpretation re-emerged in this century. I will discuss how these intellectual movements have supported the amazing diversity of applications that we see today.

I will also discuss the future. I will discuss the need to reunify the theory of probability, and how this can be done. Reunification requires, I believe, a more flexible understanding of the relation between theory and application, a flexible understanding that the decline of empiricism makes possible. I will also discuss the institutional setting for reunification: departments of statistics. Departments of statistics have been the primary vehicle for the development of statistical theory and the spread of statistical expertise during the past half-century, but they need new strategies in order to be a source of innovation in the twenty-first century. We need a broader conception of probability and a broader conception of what departments of statistics should do.

2. THE ORIGINAL UNITY OF PROBABILITY

In this section, I will sketch how the original theory of probability unified frequency, belief and fair price. (For details, see Hald, 1990; Daston, 1988; Hacking, 1975.)

In order to understand this unity, we must first understand a paradox. The original theory of probability was not about probability at all. It was about fair prices.

Probability is an ancient word. The Latin noun *probabilitas* is related to the verb *probare,* to prove. A probability is an opinion for which there are good proofs, an opinion that is well supported by authority or evidence.

Pascal and Fermat did not use the word probability in their 1654 letters. They were not thinking about probability. They were thinking about fair prices.

Here is the problem they were most concerned with, a problem that had been posed in arithmetic books for centuries, but that they were the first to solve correctly. You and I are playing a game. We have both put \$5 on the table, and we have agreed that the winner will get all \$10. The game consists of several rounds. The first player to win three rounds wins the game. I am behind at the moment—I have won one

---

Glenn Shafer is Ronald G. Harper Distinguished Professor of Business, School of Business, University of Kansas, Lawrence, Kansas 66045.

This article is based on the author's inaugural lecture as Ronald G. Harper Distinguished Professor of Business at the University of Kansas, delivered on November 20, 1989.

round, and you have won two—and I must leave to give a lecture. My wife Nell is willing to take my place in the game, taking over my position and my chance, such as it is, of winning the $10. What should she pay me for this chance? What is the fair price for my position in this game?

You have won two rounds to my one round. So perhaps you deserve two-thirds of the $10, and I deserve one-third. Pascal gave a different answer. He said you deserve three-fourths, and I deserve only one-fourth. The fair price for my position in the game is only $2.50.

Here is Pascal's argument. Were we to play the next round, we would have equal chances, and if you were to win, you would get all $10. You are entitled to $5 right there. If you were to loose, we would be even, with two games each. So we should split the other $5 equally. That leaves me with only $2.50.

Probability theory got started from this kind of reasoning. Pascal and Fermat's basic ideas were published in a short but very influential tract by the Dutch mathematician Christian Huygens. Huygens, together with the French nobleman Pierre Rémond de Montmort and the Huguenot refugee Abraham De Moivre, found fair prices for positions in more and more complicated games. The Swiss mathematician James Bernoulli even found fair prices for positions in court tennis, the complicated indoor ancestor of modern lawn tennis.

There was no talk about probability at the beginning of this work. Only equal chances and fair prices. There wasn't even a number between zero and one (my probability of winning) in the discussion. Probability was another topic. Probability was concerned with evidence, and it was a qualitative idea.

It was nearly 60 years after Pascal and Fermat's letters, in 1713, that their theory of fair price was tied up with probability. In that year, five years after James Bernoulli's death, his masterpiece *Ars Conjectandi* was published. Most of this book is about games of chance, but in Part IV, Bernoulli introduces probability. Probability is a degree of certainty, Bernoulli says, and it is related to certainty as a part is related to a whole; *Probabilitas enim est gradus certitudinus, & ab hac differt ut pars à toto.* Just as the rounds you have won and lost in a game entitle you to a definite portion of the stakes, the arguments you have found for and against an opinion entitle you to a definite portion of certainty. This portion is the opinion's probability. (Some qualifications are required here. The idea of probability was already connected to Pascal and Fermat's theory in a general way in the very influential Port Royal Logic (Arnauld and Nicole, 1662). George Hooper used the word probability to refer to a number between zero and one in work published just before 1700 (Shafer, 1986). It was the intellectual grounding provided by Bernoulli, however, that bound the idea of probability irrevocably to Pascal and Fermat's mathematics.)

Bernoulli's introduction of probability was motivated by his desire to apply the theory of fair price to problems beyond games of chance, problems *in civilibus, moralibus & oeconomicis,* problems in domains where the qualitative idea of probability had traditionally been used.

This ambition also led Bernoulli to another innovation, the theorem that is now called the law of large numbers. Bernoulli knew that in practical problems, unlike games of chance, fair prices could not be deduced from assumptions about equal chances. Chances might not be equal. Probabilities in practical problems would have to be found from observation. Bernoulli proved, within his theory, that this would be possible. He proved that if a large number of rounds are played, then the frequency with which an event happens will approximate its probability.

Bernoulli's ideas were quickly taken up by Abraham De Moivre, who made them the basis of his book, *The Doctrine of Chances* (De Moivre, 1718), which served as the standard text for probability during the eighteenth century. The French mathematician Laplace extended De Moivre's work further, into the beginning of mathematical statistics. Laplace's *Théorie analytique des probabilités* (1812), served as the standard text for advanced mathematical probability and its applications for most of the nineteenth century.

I cannot trace this development here. I do want to emphasize, however, that Bernoulli and De Moivre's mathematics bound fair price, belief and frequency tightly together. The probability of an event, in their theory, was simultaneously the degree to which we should believe it will happen and the long-run frequency with which it does happen. It is also the fair price, in shillings, say, for a gamble that will return one shilling if it does happen.

Figure 1 summarizes the logic of the classical theory. Probability (i.e., degree of certainty or degree of warranted belief) was defined in terms of fair price, and long-run frequency (or more precisely, knowledge and belief about frequencies and other aspects of the long run) was derived in turn from probability. A whole mathematical structure goes along this route; the rules for mathematical probability derive from the properties of fair price, and the details of our knowledge of the long run derive from these rules.

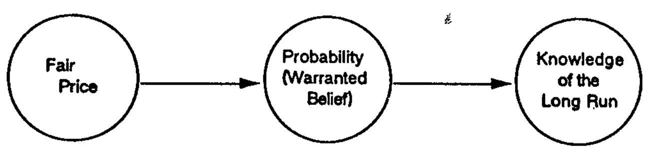

FIG. 1. *The classical theory.*

3. THE RISE OF FREQUENTISM

If you had been asked, before you began to read this article, what mathematical probability means, what would you have said?

Most people would emphasize frequency in their answer. The probability of a fair coin landing heads is one-half, they might say, because it lands heads half the time. Defining probability in terms of frequency seems to be the proper, scientific, empirical thing to do. Frequency is real. You can go out and observe it. It isn't vague, mushy and metaphysical, like "degree of certainty" or "degree of warranted belief."

If this is the way you think, then Figure 1 looks backwards to you. We should start with the facts, you will say. Start with the fact that the coin lands heads half the time. This is why even odds are fair. Don't start with some metaphysical idea about fair price and try to deduce facts from your metaphysics.

The pioneers of probability theory did not take this hardnosed empirical point of view, which seems so natural to you and me today. Our modern empiricism—positivism, it is sometimes called—is a fairly recent development in the history of ideas. It got started only in the nineteenth century.

In the case of probability, we can pinpoint just when positivism entered the stage. Independently and almost simultaneously, in 1842 and 1843, three empiricist philosophers, John Stuart Mill, Richard Leslie Ellis and Jakob Friedrick Fries, published criticisms of Laplace's classical definition of probability as degree of reasonable belief. Probability, these authors declared, only makes empirical sense if it is *defined* as frequency. So Bernoulli's theorem, which goes through mathematical contortions to prove that probability is equal to frequency, is pure nonsense. (For a fuller account of the thinking of Mill, Ellis, Fries and their allies and opponents, see Porter, 1986.)

Loath to give up Bernoulli's theorem, the mathematicians resisted this attack from the philosophers as best they could. Throughout the nineteenth century, we find probabilitists defending Laplace's ideas.

Eventually, however, probability theory came to terms with the empirical spirit of the age. There are two parts to the story of this adaptation. One part of the story is about the application of probability—how a theory of statistics was developed that was suitable for the analysis of frequency data. The other part of the story is about the mathematics of probability—how the mathematical theory was adapted to the frequentist interpretation of probability, so that you could be a frequentist and still prove Bernoulli's theorem. Both parts of the story take us through the end of the nineteenth century into the twentieth.

Statistics. In order for probability to be empirical, it should be about actual populations and actual variation in populations. But the technical content of the mathematical theory of probability in the early nineteenth century was not well adapted to the study of variation. The ideas of correlation and regression, which statisticians use nowadays to study variation, were not worked out until the end of the nineteenth century.

At the beginning of the nineteenth century, the best developed application of probability theory was in the analysis of errors of measurement, used in astronomy, geodesy and other areas of natural science. It was in this error theory that one found the normal distribution, the bell-shaped curve that statisticians now use in the study of variation. Adolphe Quetelet, the Belgian polymath who tried to apply the normal distribution to social statistics in series of publications from 1827 to 1870, ultimately failed because he was unable to escape from the conceptual setting of error theory. Just as the astronomer's measurement was an approximation to an ideal true value, Quetelet saw each individual in his human populations as an approximation to an ideal type. Quetelet's ideal was the average man.

The concepts of correlation and regression were finally discovered in the course of the study of heredity, by the Englishman Francis Galton. Galton's work was refined into a statistical methodology at the turn of the century by Karl Pearson and R. A. Fisher. All three of these scholars were genuinely interested in variation, because they were eugenicists. They were not content to regard the average Englishman as the ideal Englishman. They wanted to take advantage of variation to improve the race. (The story I sketch so briefly here is told in depth in Porter and Stigler, 1986). The influence of eugenics on the development of mathematical statistics is discussed in Mackenzie, 1981).

Once the basic ideas of correlation and regression were developed, the particular problem of eugenics faded from the center of mathematical statistics. There is variation everywhere. Yet the frequentist statistical methodology developed by Pearson and Fisher is still the core of statistical theory.

Mathematical Probability. The mathematical theory of probability was adapted to frequentism in a less direct way. The key to the adaptation was the philosophy of mathematics of the great German mathematician David Hilbert (1862-1943).

Roughly speaking, Hilbert believed that mathematics is a formal exercise, without any essential connection to reality. The business of mathematics, he held, is the derivation of formal mathematical statements—mere strings of symbols, really—from other formal mathematical statements. Getting mathematics right is a matter of making sure the derivations follow certain rules.

Hilbert's ideas inspired an effort to base all mathematics on axioms, like the axioms you learned for plane geometry in high school. Most advanced mathematics, it turns out, can be built up axiomatically starting with set theory, the abstract theory of "groups of things" invented by the German mathematician George Cantor (1845-1918). Many of the less mathematical readers of this article will remember set theory from the late 1960s and the early 1970s, when the "new math" brought it into the elementary schools in this country.

In the case of probability, the reduction to set theory was completed only in the late 1920s and early 1930s. The definitive formulation was by Andrei Kolmogorov, the great Russian mathematician who died in 1987. Kolmogorov's axioms for probability are extremely simple. They treat events as sets, and probabilities as numbers assigned to these sets, and they require that these numbers obey certain rules. The main rule is additivity. The probabilities of disjoint events add.

Kolmogorov's axioms have been extremely successful as a basis for the further development of mathematical probability. They have freed mathematicians from all the paradoxes and confusions that bedevil the interpretation of probability, leaving them a clear view of their purely mathematical problems. The axioms have been so successful, in fact, that pure mathematicians often proclaim, with a straight face, that "probability began with Kolmogorov."

I will not inflict on you the notation required to state Kolmogorov's axioms. I do want to point out, though, that these axioms take us far away from the setting in which Pascal and Fermat began, where repeated rounds of a game are played and prices and hence probabilities change. Kolmogorov's axioms are about a single probability space. Neither price nor repetition are fundamental now; they are both arbitrary elements added on top of the basic foundational axioms.

4. THE ROLE OF THE STATISTICS DEPARTMENT

Probability and statistics have become so broad that in order to understand their development in the twentieth century, we must focus on institutions rather than on individual scholars.

Though the new statistics was invented in Britain, it was taken up as a practical methodology more quickly in the United States than in Britain itself. Leadership in statistical theory, on the other hand, remained in Britain until the Second World War. American strength in statistical methodology might be attributed to our practical spirit, but it was also due to the flexible organization of American universities. (Ben-David, 1971, discusses how the departmental organization of American universities allowed the rapid development not only of statistics but also of other new fields.) American weakness in statistical theory can be attributed, paradoxically, to our relatively impractical mathematics. The American drive to match the best mathematics of Europe had led by 1900 to a dominant role for pure mathematics in American mathematics, and that dominance has persisted within our mathematics departments throughout the century. (Birkhoff, 1977, lists Thomas S. Fiske, E. H. Moore, William F. Osgood, Maxime Bôcher, and Henry Burchard Fine as the most prominent of the pure mathematicians who took over the leadership of the American Mathematical Society around 1900. For further information on the development of American mathematics, see Duren, 1989).

What were the reasons for this dominance of pure mathematics? Folklore tells us that the Americans did not feel they could compete with the Europeans in applied mathematics. Our graduate students were unwilling to spend the time needed to master both mathematics and a field of scientific application, and our universities lacked the depth in science of the European universities. We could make a mark on world mathematics only by working as far as possible from applications. (This folklore deserves serious historical examination. This would require both assessment of the possibilities in applied mathematics at the turn of the century and much archival work. In their public declarations, Fiske and his colleagues expressed strong support for applied mathematics.)

The Second World War did bring leadership in statistical theory, along with leadership in most scientific fields, to the United States. Many of the leading European theoretical statisticians, including Jerzy Neyman and Abraham Wald, immigrated to the United States before or during the war, and our military invested heavily in statistical theory. Our universities accommodated this move into statistical theory not by changing the character of their departments of mathematics but by creating departments of statistics.

The rationale for the statistics department was worked out in the late 1930s and early 1940s by a remarkable group of American statistical statesmen, including Harold Hotelling, Jerzy Neyman, W. Edwards Deming, Burton H. Camp, S. S. Wilks, Walter Bartky, Milton Friedman and Paul Hoel. It was articulated by Hotelling in two famous lectures, "The Teaching of Statistics," delivered at Dartmouth in 1940, and "The Place of Statistics in the University," delivered at Berkeley in 1946. (The written versions were published in 1940 and 1949, respectively. They were reprinted, along with comments by some of today's leaders in statistics, in Hotelling, 1988a and b.)

In Hotelling's design, the statistics department is a bridge between mathematics and the disciplines in the university that use statistical methods. This bridging role can be seen in the undergraduate curriculum of the department, in its graduate curriculum and in its faculty's research.

Most statistics departments have relatively few undergraduate majors; they play a service role at the undergraduate level, while relying on mathematics departments to train undergraduates for their own graduate programs.

The graduate curriculum is divided between mathematical probability, with at least a few courses taught at the most austere level, and statistics, with a few basic courses taught abstractly and others in a more practical spirit. Thus each doctoral student is forced to make for him or herself the journey from mathematics to applications.

The faculty for courses in mathematical probability often have joint appointments with the mathematics department; sometimes they are simply drawn from the mathematics department. More importantly for the university, the statistics department seeks joint appointments with other departments that use statistics, from electrical engineering and geology to psychology and educational research. The faculty with joint appointments in these user departments generally have degrees in statistics and regard statistics as their primary home. Their role is to transfer the latest statistical methodology to potential users. They also provide for communication in the opposite direction; by consulting in particular applied fields, statisticians develop interests in new statistical problems in those fields, and they communicate these problems, along with their own attempts at solutions, to their statistical colleagues.

Hotelling's design has been very successful. There are now over 60 statistics departments in this country, generally at the larger public and private universities. Smaller colleges cannot afford statistics departments; but they have followed the lead of the statistics departments with various joint departments and degree programs. All told, degrees in statistics are given by over 200 colleges and universities in the United States.

5. THE REVIVAL OF SUBJECTIVE PROBABILITY

Statistics departments are a product of frequentism, and the teaching in statistics departments is still predominantly frequentist in philosophy. Yet frequentist statistical theory has its difficulties and limitations, and these have become increasingly obvious with age. I cannot detail these shortcomings here, but I must point out that they have led to resurgence of subjective ideas within statistics during the past 30 years. Since the publication of L.J. Savage's *Foundations of Statistics* in 1954, a minority of statisticians (The "Bayesians") have revived the view that probability means degree of belief. The Bayesians have had a great impact not only in statistics, but also in economics, psychology, computer science, business and medicine.

The intellectual foundation for this subjectivist revival was laid earlier, in the 1920s and 1930s, by the English philosopher Frank Ramsey and the Italian actuary Bruno de Finetti. Ramsey and de Finetti saw a way to make degree of belief, as opposed to frequency, respectable within positivist philosophy. We can give degree of belief an empirical, behaviorist interpretation by insisting that people be willing to bet on their beliefs. A degree of belief of 2/3 in rain, for example, can be interpreted as a willingness to take either side of a 2-to-1 bet on rain.

The revival of the subjective interpretation was facilitated, paradoxically, by Kolmogorov's axioms. Though these axioms were meant by Kolmogorov as a mathematical foundation for the frequentist interpretation, their formality makes them equally susceptible to a subjective interpretation. Indeed, since they do not require a structure for repetition, the axioms play into de Finetti's contention that repetition is not necessary for mathematical probability to be meaningful. In Kolmogorov's framework, structures for repetition are built on top of the axioms and are therefore optional. In the new subjective theory, repetition is optional in the interpretation of the theory as well.

Within statistics, Bayesianism amounts to a minority view about how to solve the standard problem of modeling statistical variation. We just add to the class of models we are considering some prior subjective probabilities about which model is correct. But beyond statistics, Bayesianism cuts a wider swath. During the past 30 year, it has allowed probability to penetrate into areas where statistical modeling is inappropriate because statistical data is unavailable, but where evidence is sufficiently complicated to make quantitative judgments useful.

The best known practical Bayesian technique is the decision tree, which originally appeared in Abraham Wald's frequentist statistical decision theory, but which, since the late 1950s, has been used more and more with subjective probabilities. Subjective decision trees have long been a standard topic in the undergraduate business curriculum, and now they are spreading to many other fields, including medicine and engineering.

Bayesian decision theory, the abstract version of subjective decision trees, has also become influential in philosophy and psychology. Philosophers debate whether Bayesian methods constitute a standard of rationality, and psychologists study the extent to which they describe actual human behavior under uncertainty.

The greatest impact of the revival of subjective probability has come in theoretical economics. The Bayesian model of rationality has found a role in a plethora of micro-economic models during the past 25 years, and in the past 10 years it has had a growing role in macro-economics as well.

I must also mention the new and growing influence of subjective probability in artificial intelligence. Since its inception in the 1950s, this branch of computer science has seen symbolic logic as its principal mathematical tool. But in the past 10 years, the desire to build expert systems in areas where uncertainty must be explicitly managed has inspired new interest in subjective probability judgment, and new work in probability theory. This new work provides a new perspective on probability, a perspective that puts much more emphasis on the structure of conditional independence than on numbers. It has also stimulated new work on the theory of belief functions, an alternative theory of subjective probability on which I have worked for many years (see Shafer and Pearl, 1990).

6. THE BALKANIZATION OF PROBABILITY

I have been painting a picture of intellectual vitality. The mathematical theory of probability has been flourishing, spilling over all disciplinary and institutional boundaries. But this wild growth has its negative aspects. Conceptually and institutionally, probability has been balkanized.

Twenty-five years ago, the statistics department was clearly the intellectual center of probability. Those in other disciplines who wanted to use probability or statistics came to the statistics department to study these subjects. Those who were concerned about the meaning of probability came to the statistics department to hear the debate, then still fresh and stimulating, between frequentists and Bayesians. Today, the picture has changed. On both the practical and philosophical sides, many of the new developments in probability are now taking place outside the statistics department.

Within statistics, we still introduce probability theory either as a prelude to statistical modeling or as a prelude to probability as pure mathematics. "Probabilist," to us, still means mathematician. We have not adapted our teaching to serve students who want to use probability in computer science, engineering or theoretical economics. Consequently, new traditions for teaching probability are growing up within these disciplines. Whereas our approach once provided a common language for all areas of application, it is now in danger of being reduced to one voice in a tower of Babel.

On the philosophical side, the debate between frequentists and Bayesians within the statistics department has calcified into a sterile, well-rehearsed argument. The real debate has moved outside the statistics department, and the main divisions over the meaning of probability now follow disciplinary lines. Frequentists predominate in statistics and in the experimental sciences, while Bayesians predominate in the professional schools, theoretical economics, and artificial intelligence. To most Bayesians, the debate in statistics now seems parochial; it is concerned only with statistical modeling, not with the larger issues. Most frequentist statisticians, on the other hand, see Bayesians in other disciplines as cranks. Today business and engineering schools, in their brashness and practicality, may do more than statistics departments to bring together the broad range of interpretations of probability.

In addition to failing to occupy the new ground of probability, the statistics department is also losing much of the ground it did occupy. Its role as a bridge from mathematics to users of statistics in engineering and the sciences has declined over time. This is due in part to the growth of statistical expertise within these disciplines, which makes the outside specialist less needed.

Hotelling argued that students in all fields that use statistics should take their first course in statistics from the statistics department. Only the statistical specialist, he argued, would have the mathematical grasp of statistical theory needed to teach the subject well. We have clung to the element of truth in Hotelling's argument, but in most universities we have lost the argument. This is only partly because the level of mathematical competence in the other disciplines has risen. It is also because progress has changed what the disciplines want their students to learn in the introductory course. Students in the social and biological sciences who come to learn statistics now surely deserve to be taught not only the logic of the subject but also the decades-long record of its successes and failures in their discipline. If the statistics departments cannot undertake this task, the disciplines must.

The growing isolation of the statistics department is due in part to its mathematization. Perhaps any discipline that serves as a bridge between mathematics and an area of application will tend, once the leadership of its founders is gone, to move back towards mathematics. It is clear that his has happened in statistics. Today most of the articles in the leading statistics journals are so mathematical that the postwar founders of our statistics departments would not be able to read them, and so impractical that they would not want to. The joint appointments that made statistics departments so influential in the 1950s and

1960s have become difficult to replicate. When we look at the departments where these appointments were most successful, such as Stanford and Wisconsin, we find that few such appointments were made in the 1970s and 1980s. Younger statisticians have had to concentrate on their mathematics in order to be recognized as first-rate.

I want to mention one more area in which the leadership role held by statisticians through the 1970s has been wrested from us. This is in our own history. When ours was a young and brash field, we controlled our own history by default. No one else cared. We didn't really know much about this history, but we told what we knew with authority. This, too, has changed. Starting with the philosopher Ian Hacking's book *The Emergence of Probability* in 1975, we have seen our history taken over by philosophers and professional historians of science. In the past decade, we have seen more books and articles on the history of probability and statistics than were published during the entire preceding existence of these subjects. The more technical of these works still tend to be written by statisticians, but most of the books that try to describe the big picture are now written by historians with relatively little technical training. The story they tell sounds very different from the story we once told. Whereas we saw probability and its progress from the inside, the historians see probability and its vicissitudes as the result of larger cultural and historical forces.

I do not believe that the balkanization of probability is a good thing. We need ways of understanding the unity that still exists in probability, and we need an institutional center for probability and statistics. We need institutions that can bring together divergent tendencies in philosophy and application, so that these tendencies can learn from each other.

In the remainder of this article, I will be concerned with how probability can be reunified. On the conceptual side, I will sketch how we can recreate an understanding of probability that has room for both frequentist and Bayesian applications, without the worn-out dogmas of either group, and also room for the newer applications. On the institutional side, I will advance some theses about what our statistics departments must do to recover their leadership role.

7. THE CONCEPTUAL REUNIFICATION OF PROBABILITY

On the conceptual side, I believe we can go back to the original unity of belief and frequency. The positivism that drove these two aspects apart no longer holds quite the sway that it held in the nineteenth century. For the positivists, every element of a theory—every object in the theory and every relation between objects—had to have a definite, verifiable, empirical reference. Today, it is possible to be more flexible about the relation between a theory and its application.

The point is that we can apply a mathematical theory to a practical problem even though it does not model that problem empirically. In order to apply a theory to a problem, it is sufficient that we relate the problem, perhaps even very indirectly, to another problem or situation that the theory does model.

The only situation that the mathematical theory of probability models directly is still the very special situation studied by Pascal and Fermat, the special situation where we flip a fair coin or play some other game with known chances. In this special situation, which I call the *ideal picture of probability,* the unity of belief and frequency is unproblematic. If we know the frequency with which a coin lands heads, this known frequency is a sensible measure of the degree to which we should believe it will land heads on any particular flip.

Whenever we use mathematical probability in a practical problem, we are relating that problem, in one way or another, to the ideal picture. In my forthcoming book, *The Unity and Diversity of Probability,* I argue that the different ways Bayesians, frequentist and others use probability should be thought of as different ways of relating problems to the ideal picture. Much standard statistical modeling amounts to using the ideal picture as a standard of comparison. Statistical arguments based on sampling or experimental randomization depend on artificially generated random numbers which simulate the ideal picture, and they relate this simulation to real problems in clever ways. Bayesian arguments can be thought of as arguments by direct analogy to the ideal picture. And arguments based on the theory of belief functions involve analogies that are less direct.

Recognition of this continuing centrality of the ideal picture will allow us to move back to unified understanding that we find in Pascal, Fermat, Bernoulli and De Moivre. We can insist on the unity of belief and frequency in the ideal picture even while admitting that they go their separate ways in many applications.

We cannot simply return to the mathematics of the seventeenth and eighteenth century, for we have learned much since that time. We can, however, reformulate the mathematical foundations of probability in a way that incorporates rather than ignores the pioneers' insights into the fundamental role of fair price and repetition. In *The Unity and Diversity of Probability,* I argue for reformulating Kolmogorov's axioms in the framework of a sequence of experiments, in which the mathematics of Figure 1 can be recaptured.

We do need to go beyond Figure 1 in one important respect. Work in the twentieth century by the frequentist scholars Richard von Mises, Jean Ville and Abraham Wald has shown that the theory of probability can be developed mathematically starting with the knowledge of the long run, which includes both knowledge of long-run frequency and knowledge of the impossibility of gambling schemes (see Martin-Löf, 1969; Cover, Gacs and Gray, 1989). Thus Figure 1 can be expanded to Figure 2, which shows fair price, belief and frequency bound together in a triangle. From a purely mathematical point of view, any point in this triangle can be taken as an axiomatic starting point, but from a conceptual point of view, none of these starting points can stand on its own. The axioms or assumptions that we must set down when we start from any one of the starting points can be justified only by reference to the other ideas in the triangle (see Shafer, 1990b).

8. THE INSTITUTIONAL REUNIFICATION OF PROBABILITY

What can be done to make the statistics department once again the intellectual center of probability?

The well-worn answer is that we should try harder to live up to Hotelling's design. We should teach better, so that other departments will send their students to us rather than developing their own probability and statistics courses. We should play university politics better, so that these departments are not allowed to develop their own courses. We should examine the mathematics we do more critically, to make

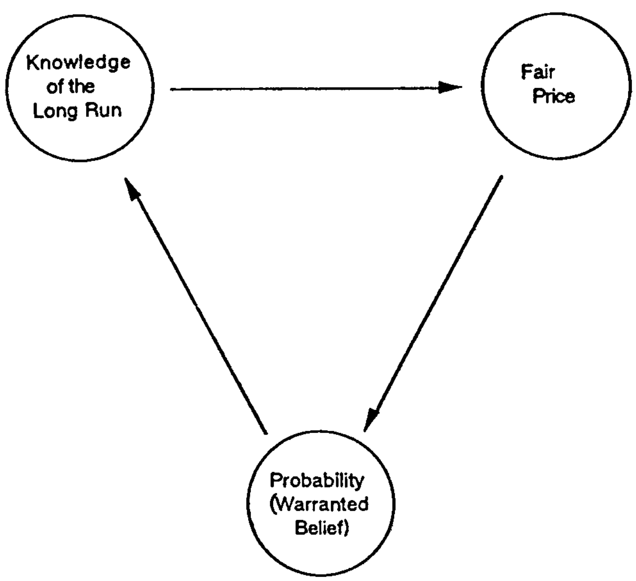

FIG. 2. *The theory of probability starting with knowledge of the long run.*

certain it is relevant to applications. We should produce more statisticians who are so bright that they can meet today's standard for mathematical accomplishment while continuing the tradition of being involved in practical problems.

Our best statistics departments are doing these things. This has not been enough, however, to stop the balkanization I have described. I believe the time has come to address the problem directly. We need a new conception of the statistics department, one that suits our times.

The new statistics department should assess and absorb into its teaching and research what other disciplines have learned about probability and statistics. The department's introductory teaching of probability, at both undergraduate and graduate levels, should be comprehensive enough to serve all users. The department should teach not only the logic of statistics, but also the issues involved in its applications. In our undergraduate statistics courses, we should try to assess past performance and future prospects for statistics in each of the disciplines we serve. Our graduate teaching should include comparative assessment of the possibilities for statistics in different fields, including both the disciplines we serve within the university and fields that we serve outside the university, such as the census.

Our best statisticians have been willing to evaluate the uses to which other fields put statistical ideas. (See, for example, the debate between John Tukey and the economists Heckman and Robb in Wainer 1986). The discipline of statistics has failed however, to produce broad assessments of the use of statistics in different fields, with examples of successes as well as failures, and with lessons for practitioners. Such assessments should be a major research and teaching goal for the new statistics department.

Where can we find the faculty for these tasks? We do not have to look far. We must co-opt the historians, computer scientists, philosophers, economists, psychologists and others who are contributing to our understanding of probability and statistics. We must enlist faculty from these disciplines to help us in our teaching mission, at both the undergraduate and graduate levels. We must make dissertations and careers concerned with aspects of probability and statistics that have been developed in these other disciplines possible within statistics.

In some cases, we should recruit people trained in these disciplines as full-time members of the statistics department. In other cases, we should ask them, in their role as faculty members in another department, to serve on an advisory committee for statistics and teach courses in the statistics department. In other cases, we should seek joint appointments. In the past, we have thought of joint appointments as a way for statistics to contribute to other disciplines. We must now think of them also as a way for the other disciplines to contribute to statistics.

Computer science is one of the first disciplines with which we should seek joint appointments. Probability has begun to play a whole spectrum of roles in computer science, from a tool in the evaluation of algorithms to a model for distributed processing to a model for learning and inference in artificial intelligence. This, together with the ever increasing role of computing in both theoretical and applied statistics, makes it essential that ties between statistics and computer science be cultivated.

History is another field with which we need reciprocal ties. Historians need our help, for in recent decades they have joined the social sciences as users of statistics. We need their help in order to carry out the assessments of statistical practice that I am advocating. Our task is to write the history of probability and statistics in the twentieth century.

I subscribe to David S. Moore's thesis that statistics belongs among the liberal arts (Moore, 1988). I believe, moreover, that we cannot teach statistics as a liberal art unless we practice it as a liberal art. The research and graduate program in the statistics department should include real attention to the history and philosophy of probability and statistics.

The proposals I have just made are far-reaching. Their implementation will not be easy or painless. It will take many years to reshape our curriculum in the directions I have suggested, and when this has been accomplished, we will have to deal with much more diverse colleagues and students than we have dealt with in the past. Evaluation of students and faculty will be more difficult and possibly more contentious. The new statistics department will not work without leadership.

Is is necessary to take so difficult a path? Many statisticians do not share my conviction that the survival of the statistics department is threatened by the balkanization of statistics. They are willing to cede the new applications of probability and the more mundane topics of applied statistics to other departments, confident that the statistics department will remain indispensable as a home for those at the forefront of research in mathematical statistics. The need for this research seems to guarantee the survival of the statistics departments.

I believe this is true in the short run. But those who would rely on the prowess of a mathematical elite for the survival of statistics as a separate discipline should look over their other shoulder. Since the David Report in 1984 (Committee on Resources for the Mathematical Sciences, 1984), the mathematics community in this country has taken remarkable strides in broadening its conception of its subject. Incredible as it seems to those of us who studied mathematics in the heyday of American fascination with the mythical French pure mathematician Bourbaki, many American mathematicians are now broadly interested in applications. It is conceivable that in the next generation we will see mathematics departments capable of interacting with a broad range of disciplines. Were this to happen, the independence of elite departments of very mathematical statisticians would no longer make sense. Such departments would be reabsorbed into mathematics.

As stewards of a legacy from a line of giants stretching from Pascal to Hotelling, we should not relish such an outcome. Statistics, as a discipline, has proven fruitful because it has had an intellectual basis broader than mathematics. Because it has been rooted in the practical and philosophical problems of inference as well as in mathematics, statistics has been able to play a leadership role extending throughout the sciences. Our goal today should be a renewal of statistics that will keep it in this position of leadership.

ACKNOWLEDGMENT

In revising the lecture for publication, the author has profited from comments from Paul Benacerraf, Theodore Chang, Arthur P. Dempster, Lorraine Daston, Nell Irvin Painter, Theodore M. Porter, G. Baley Price, Stephen M. Stigler, Geoffrey S. Watson, and Sandy L. Zabell. Research for the lecture and article was partially supported by Grant IRI-8902444 from the National Science Foundation.

REFERENCES

ARNAULD, A. and NICOLE, P. (1662). *l'Art de Penser.* Paris.

BEN-DAVID, J. (1971). *The Scientist's Role in Society: A Comparative Study.* Prentice-Hall, New York.

BERNOULLI, J. (1713). *Ars conjectandi.* Basel.

BIRKHOFF, G. (1977). Some leaders in American mathematics: 1891–1941. In *The Bicentennial Tribute to American Mathematics, 1776–1976* (D. Tarwater, ed.) 25–78. Mathematical Association of America, Washington, D.C.

COMMITTEE ON RESOURCES FOR THE MATHEMATICAL SCIENCES (1984). Renewing U.S. mathematics: Critical resources for the future. *Notices of the American Mathematical Society* **31** 434–466.

COVER, T. M., GACS, P. and GRAY, R. M. (1989). Kolmogorov's contributions to information theory and algorithmic complexity. *Ann. Probab.* **17** 840–865.

DASTON, L. (1988). *Classical Probability in the Enlightenment.* Princeton Univ. Press.

DE MOIVRE, A. (1718). *The Doctrine of Chances.* London. (Later editions 1738, 1756.)

DUREN, P. (ed.) (1989). *A Century of Mathematics in America.* 3 volumes. Amer. Math. Soc., Providence, R.I.

HACKING, I. (1975). *The Emergence of Probability.* Cambridge Univ. Press, London.
HALD, A. (1990). *A History of Probability and Statistics and their Applications before 1750.* Wiley, New York.
HOTELLING, H. (1988a). The teaching of statistics. *Statist. Sci.* **3** 63–71.
HOTELLING, H. (1988b). The place of statistics in the university. *Statist. Sci.* **3** 72–83.
LAPLACE (1812). *Théorie analytique des probabilités.* Paris. (Later editions 1814, 1820.)
MACKENZIE, D. A. (1981). *Statistics in Britain, 1865-1930: The Social Construction of Scientific Knowledge.* Edinburgh University Press.
MARTIN-LÖF, P. (1969). The literature on von Mises' Kollektivs Revisited. *Theoria* **35** 12–37.
MOORE, D. S. (1988). Comment on "The teaching of statistics" and "The place of statistics in the university." *Statist. Sci.* **3** 84–87.
PORTER, T. M. (1986). *The Rise of Statistical Thinking.* Princeton Univ. Press.
SAVAGE, L. J. (1954). *Foundations of Statistics.* Wiley, New York.
SHAFER, G. (1986). The combination of evidence. *Internat. J. Intelligent Systems* **1** 127–135.
SHAFER, G. (1990a). *The Unity and Diversity of Probability.* In preparation.
SHAFER, G. (1990b). The unity of probability. In *Acting under Uncertainty: Multidisciplinary Conceptions* (G. M. von Furstenburg, ed.) 95–126. Kluwer, New York.
SHAFER, G. and PEARL, J. (eds.) (1990). *Readings in Uncertain Reasoning.* Morgan-Kaufmann, San Mateo, Calif.
STIGLER, S. M. (1986). *The History of Statistics.* Harvard Univ. Press, Cambridge, Mass.
WAINER, H. (ed.) (1986). *Drawing Inferences from Self-Selected Samples.* Springer, New York.

Comment

Hirotugu Akaike

Professor Shafer's paper shows his concern for the future of statistics. He considers that the present situation of statistics is alarming and, assuming that mathematical statistics is a child of mathematical probability, attributes this situation to the popularization and diversification of the use of probability. I completely agree with Professor Shafer on the recognition of the problematical status of statistics and would like to add some observations on the nature of statistics and probability.

STATISTICS FOR PLANNING AND PROBABILITY FOR DECISION

It is almost certain that the original concept of statistics started with the description of the state of a nation by counting and classifying its people. Any country appearing in the history must have used some kind of statistics for the management of the country. Along with this very old origin of the concept of statistics was also the use of probabilistic mechanisms or randomizers by ancient kings.

A typical example of the use of a randomizer is given by the *I Ching*, or the Book of Changes, which shows the wisdom of ancient Chinese people for the handling of uncertainties. With this book there is an advice that recommends the minimum use of the book to attain a proper objective.

Hirotugu Akaike is Director General of the Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan.

Consider a king who was going to declare a war against another country. It is almost certain that he used statistics for the planning of the war. But there must have remained some uncertainty. If he intended to consult the *I Ching*, then the advice would have forced him to make utmost effort to minimize the uncertainty before he turned to the randomizer. This means that the process of setting up a probability distribution for a particular purpose must be based on a fully efficient use of available information which is often supplied by the related statistical data in the case of the decision related to the future of a nation.

Here we can see a typical example of the use of statistics for planning and probability for decision. This example also demonstrates the inherent connection of probability and statistics with the proper use of information.

PROBABILITY OF A SINGLE EVENT

Consider a situation where probability $p(A)$ of the occurrence of an event A is given. When $p(A)$ is greater than 0.5, according to the interpretation of probability as described by Shafer, it would seem reasonable to bet on the occurrence of A. However, since probability does not tell anything about actual occurrence of a particular event, some justification is required for the decision to bet on A.

This problem is deeply related to the argument of objectivity or subjectivity of probability. If the probability is considered to be objective, in the sense that it is accepted by most of the members of a society, the