International Statistical Review, (1985), 53, 3, pp. 261–277. Printed in Great Britain

Conditional Probability

Glenn Shafer

School of Business, University of Kansas, Lawrence, Kansas 66045, USA

Summary

Expected values in a game of chance change with the step-by-step unfolding of the game. This unfolding is governed by a protocol, a set of rules that tell, at each step, what can happen next. This paper develops the idea of a protocol intuitively, historically and mathematically. Protocols are important in statistics and in subjective probability judgment because we can properly interpret new information only when we know the rules governing its acquisition. With a protocol, the rule of conditioning can be treated as a theorem. Without a protocol, the use of this rule is questionable.

Key words: Missing data; Protocol; Selection; Subjective probability.

1 Introduction

Mathematical probability began with the picture of games of chance. This picture is based on the idea that gambles have fair prices, and it emphasizes that these fair prices change with the step-by-step unfolding of the game. The unfolding is governed by a protocol, a set of rules that tell, at each step, what can happen next.

During the nineteenth and early twentieth centuries, the idea of fair price lost its dominant role in mathematical probability. In its place, two competing conceptions of probability came to the fore: the conception of probability as relative frequency, and the conception of probability as rational degree of belief. Under both these conceptions, the idea of a protocol for changes in probability has tended to disappear. We do not seem to need a protocol telling us how we may change the reference class for our relative frequencies or the evidence for our rational degrees of belief.

The mid-twentieth century has seen a further conceptual evolution. Today, most scholars who understand probability as degree of belief have dropped the adjective 'rational'. These scholars no longer hold, as Keynes did, that given evidence logically determines a probability for a given proposition. Instead, they think of a person's probability for a proposition as his personal degree of belief: personal in that it depends on his idiosyncrasies as well as on his information. Since this personal degree of belief is defined behavioristically, in terms of choices the person might make, this mid-twentieth century evolution can be seen as a partial return to the original picture of games of chance. A probability is again a price for a gamble, a price a person considers fair inasmuch as it is the price at which he will both buy and sell.

In this paper I argue that we need to complete this evolution back to the original picture of games of chance by reintroducing the notion of a protocol. I argue that the Bayesian picture of subjective probability, the picture created by Bruno de Finetti and L.J. Savage, becomes, when it is fully developed, the picture of a person who not only has prices for uncertain rewards but also has a protocol laying out the possibilities for what new information or evidence he may acquire.

The de Finetti–Savage picture shows us a person who has a subjective probability distribution Pr over a set of possibilities Ω. When the person acquires new information, he recognizes that this information tells him that the truth is in a certain subset B of Ω, and he changes his probability for any given subset A of Ω from the unconditional probability $\Pr[A]$ to the conditional probability $\Pr[A \mid B]$, where

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}. \tag{1}$$

As it is usually presented, this picture seems to suggest that the person may be totally uninformed about what 'new information' may come along and hence may be prepared to apply (1) to any subset B of Ω, at least any subset B such that $\Pr[B] > 0$. Here I argue that this is wrong. The change represented by (1) is defensible and justifiable only on the basis of a protocol that tells circumstances under which the information B will be acquired. And any such protocol will justify (1) only for some subsets B of the full set of possibilities Ω.

This conclusion has important implications for subjective probability. It makes clear how thoroughly the Bayesian theory assimilates practical problems to the picture of games of chance, and it forces us to ask how this assimilation can be justified. How can a person be expected to have a protocol for new information? If he cannot be expected to do so, do we not need a new understanding of how the de Finetti–Savage picture is to be related to practical probability judgment?

We do not always have protocols for new information in practical problems. This does not mean that we must despair of subjective probability judgment in such problems; often such judgment is necessary. But it does mean that the logic of subjective judgment is not as clear and powerful in these problems as we might like. In particular, it means that it is not normative to use the Bayesian theory in such problems. Alongside the Bayesian theory, which draws an imperfect analogy between the practical problem and a game of chance, we should consider other theories, which rely on analogies to other canonical examples (Shafer, 1985; Shafer & Tversky, 1985).

In § 2 of the present paper, I discuss intuitively what a protocol is and why it is needed. In § 3, I give a simple formalization of the notion of a protocol and use this formalization to demonstrate how (1) can be justified when there is a protocol. This demonstration, it turns out, dates back to the second edition of De Moivre's *The Doctrine of Chances*, published in 1738. In § 4 I discuss the practical and philosophical implications of the need for protocols.

There are two appendices. Appendix 1 lays the foundations for a mathematical treatment of probability and expectation with protocols. Appendix 2 relates the ideas developed in the paper to de Finetti's treatment of conditional probability.

2 Protocols for new information

When should you change your probability for A from $\Pr[A]$ to $\Pr[A \mid B] = \Pr[A \cap B]/\Pr[B]$? Everyone agrees that two conditions should be met. First, you must have $\Pr[B] > 0$. Secondly, the fact of B's happening must be the only thing you have learned. (This is the 'principle of total evidence'.) The thesis of this paper is that a third condition must be met: your learning about B's happening must be in accordance with a protocol that told you the possibilities for what you would learn.

The purpose of this section is to convey an intuitive understanding of this need for a protocol for new information. We begin by studying a puzzle that illustrates the need. We then notice that this need is implicit in the principle of total evidence.

Freund's puzzle of the two aces. This is one of many well-known puzzles that illustrate the need for protocols in conditioning. It is presented here in a form close to the

form in which it was presented by John E. Freund in *The American Statistician* in 1965. The resolution given here is in agreement with Schrödinger (1947), Gridgeman (1963) and Faber (1976).

I show you a deck containing only four cards: the ace and deuce of spades, and the ace and deuce of hearts. I shuffle them, deal myself two of the cards, and look at them, taking care that you do not see which they are. You realize that there are six equally likely possibilities:

A spades and A hearts,	A hearts and 2 spades,
A spades and 2 spades,	A hearts and 2 hearts,
A spades and 2 hearts,	2 hearts and 2 spades.

If A denotes the event that I have two aces, B_1 denotes the event that I have at least one ace, and B_2 denotes the event that I have the ace of spades, then your initial probabilities are

$$\Pr[B_1] = \tfrac{5}{6}, \quad \Pr[B_2] = \tfrac{1}{2}, \quad \Pr[A] = \Pr[A \cap B_1] = \Pr[A \cap B_2] = \tfrac{1}{6}.$$

Now I smile and say, 'I have an ace'. The usual doctrine is that you should react to this new information by conditioning on the event B_1. Thus your new probability for A, the event that I have two aces, is

$$\Pr[A \mid B_1] = \frac{\Pr[A \cap B_1]}{\Pr[B_1]} = \frac{1/6}{5/6} = \frac{1}{5}.$$

The information that I have at least one ace has increased your probability that I have two.

Now I smile again and announce, 'As a matter of fact, I have the ace of spades'. You are supposed to condition again, this time on the event B_2, obtaining the new probability

$$\Pr[A \mid B_1 \cap B_2] = \Pr[A \mid B_2] = \frac{\Pr[A \cap B_2]}{\Pr[B_2]} = \frac{1/6}{1/2} = \frac{1}{3}.$$

The more specific information that I have the ace of spades has increased even further your probability that I have two aces.

Is this second change in your probability for A reasonable? Should my decision to identify a suit make any difference?

You are puzzled about whether or not you should change your probability for A because we had not agreed to a protocol for what information I would communicate. Under some protocols, the change is reasonable; under others it is not. If it had been agreed beforehand that I would tell whether or not I had at least one ace and then whether or not I had the ace of spades, then the second step would involve relevant information. (Had I said no, I do not have the ace of spades, then your probability for my having two aces would have gone down to zero, and so it is reasonable that when I say yes this probability should go up.) On the other hand, if it had been agreed that I would first tell whether or not I had at least one ace and then, if I did have one, that I would name the suit of one I had, then the second step would not involve relevant information.

What can you do when no protocol has been agreed to? You can, if you wish, refuse to make any probability judgment at all, on the grounds that you have no way of knowing what I am up to. Alternatively, you can rely on your own, purely subjective, protocol for what information I might communicate. You can claim to have had prior probabilities concerning what, under different circumstances, I would say, and using these prior probabilities you can calculate probabilities conditional not only on the fact that I have the ace of spades, but also on the fact that I have said so, with a smile.

Note that the puzzle of the two aces appears, in various forms, in many books of mathematical puzzles, including Ball (1911, p 32), Gamow & Stern (1958, pp. 37–42) and Gardner (1959, pp. 49–51). It has also been discussed by Schrödinger (1947) and Gridgeman (1963). Freund's discussion of the puzzle in *The American Statistician* was followed by a long series of letters to the editor of that journal (Freund, 1965). The writers of these letters disagreed sharply about how to explain the puzzle. A similarly confused controversy, inspired by the presentation of the puzzle by Copi (1968, p. 433), has appeared in *Philosophy of Science*: Rose (1972), Dale (1974), Faber (1976), Goldberg (1976). Feller's puzzle of the two boys (Feller, 1950, Ch. 5, § 1) and Mosteller's puzzle of the three prisoners (Mosteller, 1965, problem 13), which turn on the same point, have also inspired much comment. Bar-Hillel & Falk (1982) discuss these puzzles from a psychological point of view. Similar puzzles arise in actual card games; see, for example, Rubens (1976).

The principle of total evidence. The need for a protocol is implicit in the principle that you should condition on everything you learn.

To see this, notice that when you learn something, you also learn that you have learned it. (When I tell you I have the ace of spades, you learn not only that I have it, but also that I told you so.) So the event B that specifies all the facts you learn must also specify the fact that you learn these facts. Indeed, B is most succintly described as the event that the totality of what you learn is such-and-such. (Once we have said you learned a fact, it adds nothing to say it is a fact.) But when we say you should condition on B, we are assuming B is in your probability model, i.e. in the field of events to which you assign probabilities. So it is implicit in the principle of total evidence that your probability model should include a model for what you learn. It should give probabilities for the different possible ways your learning may turn out. This amounts to giving a protocol for new information.

3 Justifying conditional probability

We can gain further insight into the nature of a protocol for new information by considering what assumptions are needed in order to derive equation (1), the *rule of conditioning*, as a theorem.

In this section I present two versions of an argument that does produce (1) as a theorem. Both versions concern subjective probabilities. And both rely on the existence of a protocol, a model for the way your learning may turn out. They differ in that the first assumes that this protocol is objective and publicly known, while the second assumes only that it gives your personal opinions about how your learning may turn out.

After presenting these two arguments, I briefly discuss the similar arguments given in the eighteenth century by De Moivre and Bayes. I postpone to Appendix 2 discussion of an argument by de Finetti which also seems similar but which does not rely on a protocol.

In order to avoid constant reference to the condition '$\Pr[B]>0$', I shall argue not for (1), but for the nearly equivalent equation

$$\Pr[A\cap B]=\Pr[B]\Pr[A \mid B], \qquad (2)$$

which necessarily holds if $\Pr[B]=0$. This equation is often called the *rule of compound probability*.

An objective protocol. Suppose it is arranged and publicly known that (i) if B happens its happening will be known immediately to you and others, and (ii) the fact of its happening will be the only new information you will receive up to that point. Suppose, further, that your probabilities, now and later, are interpreted as two-sided betting rates. When you announce a probability p for an event, you are offering to bet for or against the event at odds $p:(1-p)$.

Since B's happening is the only new fact you will learn by the time of its happening, your total evidence at that time is well defined. So you can think about what your probability for another event A will be right after B happens (if it happens). Let $\Pr[A \mid B]$ denote the number you now think this probability will be.

Suppose you are required to annouce both $\Pr[A \mid B]$ and your present probabilities $\Pr[A\cap B]$ and $\Pr[B]$. Then following Teller (1973), we can show that these probabilities must obey (2) or else an opponent can construct from your offers to bet a 'Dutch book', an arrangement whereby you are certain to suffer a net loss.

Suppose, indeed, that (2) does not hold. Consider the case where

$$\Pr[A\cap B]>\Pr[B]\Pr[A \mid B].$$

Suppose an opponent makes the following bets with you.

(i) He bets against $A\cap B$, with total stakes of \$1. This means he will pay you $\$(1-\Pr[A\cap B])$ if A and B both happen and otherwise will collect $\$\Pr[A\cap B]$ from you.

(ii) He bets on B, with total stakes of $\$\Pr[A \mid B]$. This means he will collect $\$(1-\Pr[B])\Pr[A \mid B]$ from you if B happens and otherwise will pay you $\$\Pr[B]\Pr[A \mid B]$.

(iii) The third bet he makes only if B happens. If B does happen, he then bets on A, at your new odds, with total stakes of \$1. This means he will collect $\$(1-\Pr[A \mid B])$ from you if A and B both happen, and he will pay you $\$\Pr[A \mid B]$ if B happens but A does not.

The amounts the opponent will collect are shown in Table 1. Negative entries are payments he makes to you. As the totals of the three columns indicate, he will have a net gain of $\$(\Pr[A\cap B]-\Pr[B]\Pr[A \mid B])$ no matter what happens. By assumption this net gain is positive.

If $\Pr[A\cap B]<\Pr[B]\Pr[A \mid B]$, the opponent uses the same bets but switches sides, thus reversing all the signs in Table 1.

Notice that the argument depends not only on the protocol for what you are going to learn but also on the requirement that you announce beforehand and follow a plan for changing your probabilities. An opponent cannot make a Dutch book against you if he does not know how you are going to change your probabilities, but he can make a Dutch book against you if you follow an announced policy that differs from the rule of conditioning.

Table 1
The Dutch-book argument

Opponent's bets	Opponent's gains: If A and B both happen	If B happens but A does not	If B does not happen
(i) Against $A\cap B$	$-(1-\Pr[A\cap B])$	$\Pr[A\cap B]$	$\Pr[A\cap B]$
(ii) On B	$(1-\Pr[B])\Pr[A \mid B]$	$(1-\Pr[B])\Pr[A \mid B]$	$-\Pr[B]\Pr[A \mid B]$
(iii) On A (if B happens)	$1-\Pr[A \mid B]$	$-\Pr[A \mid B]$	—
Totals	$\Pr[A\cap B]-\Pr[B]\Pr[A \mid B]$	$\Pr[A\cap B]-\Pr[B]\Pr[A \mid B]$	$\Pr[A\cap B]-\Pr[B]\Pr[A \mid B]$

The crux of the argument is the fact that bets (i) and (ii) together amount to a bet on the event $A \cap B$. If you announce $\Pr[B]:(1-\Pr[B])$ as your odds for betting for or against B and you also announce $\Pr[A \mid B]:(1-\Pr[A \mid B])$ as the odds at which you will bet for or against A once you have learned of B's happening, then an opponent can arrange what amounts to a bet on $A \cap B$ at odds $\Pr[B]\Pr[A \mid B]:(1-\Pr[B]\Pr[A \mid B])$ by betting on B at total stakes $\$\Pr[A \mid B]$ and arranging to bet on A at total stakes \$1 if and when you learn B has happened. Hence you have in effect announced a probability of $\Pr[B]\Pr[A \mid B]$ for the event $A \cap B$.

A subjective protocol. Consider now the case of a purely subjective protocol. Suppose this protocol says (i.e., you believe) that if B happens its happening will be immediately known to you and will be your only new information by that time. But this is not a matter of public knowledge. There is not necessarily an opponent who knows or agrees that B has this special property. Suppose again that there is a definite number, which we denote by $\Pr[A \mid B]$, that you think will be your probability for A if and when B happens.

In this case we cannot raise the specter of an opponent who is allowed to exploit $\Pr[A \cap B]$, $\Pr[B]$ and $\Pr[A \mid B]$ as offers to bet. We can, though, talk about what a contract is worth to you. And we can relate your valuation of risky contracts to your probabilities in the usual way: a contract that pays \$1 if an event happens is worth $\$p$ to you if your probability for the event is p.

Here is a more subtle point. Since you have opinions about what your new probabilities will be once B happens, you may also have opinions about your valuations at that point. And this, in turn, may say something about what certain contracts are worth to you now. Let us assume that a contract that pays a thing T if B happens is worth $\$\Pr[B]v$ to you now if you expect T to be worth $\$v$ to you once B has happened.

Now consider a nested contract: a contract that pays, if B happens, a contract that pays \$1 if A also happens. By the assumptions we have made, this nested contract is worth $\$\Pr[B]\Pr[A \mid B]$ to you now. But this nested contract boils down to a contract that pays \$1 if the event $A \cap B$ happens. So we have shown that $\Pr[B]\Pr[A \mid B]$ is equal to $\Pr[A \cap B]$.

It seems reasonable to say that we have just given two versions of the same argument. In the case of the objective protocol, we combine bets you are willing to make to show that you would pay (or accept) $\$\Pr[B]\Pr[A \mid B]$ for a bet that would return \$1 if A and B both happened. In the case of the purely subjective protocol, we combine your valuations of risky contracts to show that a contract that will return \$1 if A and B both happen is worth $\$\Pr[B]\Pr[A \mid B]$ to you.

There is, in a sense, an escape clause in both versions. In the case of the objective protocol, you might, when B happens, violate your announced policy and adopt a probability for A that differs from $\Pr[A \cap B]/\Pr[B]$. In the case of the subjective protocol, you might, when B happens, violate your own expectations and adopt a probability for A that differs from $\Pr[A \cap B]/\Pr[B]$. We have shown not that you will change your probabilities in conformity with the rule of conditioning, but merely that you expect to do so.

De Moivre and Bayes. The eighteenth-century writers Abraham De Moivre and Thomas Bayes gave their own versions of our argument for the rule of compound probability.

De Moivre's argument first appears on pages 5–6 of the second edition of his book, *The Doctrine of Chances*, published in 1738. (It is repeated unchanged in the third edition, published posthumously in 1756.) He spells the argument out only for one particular numerical example of independent events. But this is a matter of exposition; he clearly intends the reader to extend his argument to the general case, where the events may be dependent. His argument is very similar to the second of our two versions, for it is based on the assumption that an expectation, or a contract that returns a certain sum of money if a specified event happens, has a definite value, proportional to the probability of the event.

It would be misleading, however, to call De Moivre's argument subjective. He seems not to have distinguished between objective and subjective protocols, or even between objective and subjective probabilities. He was directly concerned with games of chance, where probabilities, values of expectations, and protocols are determined by publicly known rules and hence are objective as well as subjective.

We find another version of De Moivre's argument in the proof of Proposition 3 of Thomas Bayes's famous essay on probability, which was published posthumously in 1764. Bayes was also concerned with a question that had not occurred to De Moivre. He was interested in the probability we should give to a one event given knowledge only of the happening of a later event. Though Bayes did not explicitly distinguish objective and subjective probabilities, his question clearly leads to such a distinction. If our knowledge does not keep up with the objective unfolding of events, then there is a clear sense in which our probabilities are only subjective. Indeed, we can imagine the existence, in the same game of chance, of distinct objective and subjective protocols. The objective protocol would govern the actual happening of events, and the subjective protocol would govern our learning of them. De Moivre's argument could be applied to both protocols, determining objective and subjective probabilities that diverge as events proceed. Bayes did not, however, formulate the idea of a subjective protocol. Instead he attempted, in his Proposition 5, to justify the rule of conditioning without a subjective protocol. This attempt was not successful (Shafer, 1982).

4 Should protocols be part of the theory of probability?

Freund's puzzle shows that a protocol is sometimes needed to make conditioning on new information legitimate. And we have seen that this legitimatization can be built into the Bayesian theory of subjective probability. Should it be built in? Should we make the role of protocols explicit and treat the rule of compound probability as a theorem within the theory? Or should we leave protocols outside the theory, in the realm of informal judgment that governs our use of the theory?

Almost all twentieth-century writers on subjective probability have left protocols in the realm of informal judgment. This is due, in part, to the influence of A.N. Kolmogorov's axiomatization of probability, in which (1), the rule of conditioning, is interpreted as a definition of the conditional probability $\Pr[A \mid B]$, a definition that is equally meaningful for all B such that $\Pr[B]>0$. As soon as we call the rule of conditioning a definition, we have pushed the question of when the rule should be used outside our theory.

Much is to be gained by bringing protocols back into the theory of subjective probability. I do not mean to denigrate the historical importance of Kolmogorov's axioms nor to deny that experienced and careful statisticians, working on specific concrete problems, can make wise informal judgments about whether conditioning is justified. But I believe an explicit recognition of the need for protocols can bring many insights to those who are less experienced or who undertake to think about subjective probability in abstraction from specific problems.

Insights in applied statistics. Statistical theory has traditionally been based on the idea of a planned experiment. The set of possible outcomes of the experiment, known in advance, is called the *sample space*. From the Bayesian viewpoint, the sample space amounts to a protocol for new information. But because its role as a protocol is only implicit, we can easily lose sight of this role when statistical ideas are applied outside the realm of planned experiments.

People often fail to realize, for example, that the significance of a surprising fact depends on the search that turned it up. Consider the three elderly people with cardiac problems who died within a few hours of being vaccinated for swine flu in Pittsburgh in 1976. Many people were puzzled when statisticians said that the significance of these deaths depended on something poorly known: how sharp an eye was being kept out for such coincidences (Kac & Rubinow, 1977; Neustadt & Fineberg, 1982). Similar problems arise when the results of medical experiments are scrutinized for apparent effects for subgroups of the population being studied (Tukey, 1977). It is also easy, both in experiments and in surveys, to overlook the need for a protocol governing whether data will be missing (Dawid & Dickey, 1977; Rubin, 1976, 1978.)

It is possible to deal with all these examples within a framework where the rule of conditioning is called a definition. We simply insist, in each case, that the 'correct sample space' be used; namely, a sample space that models the process of data acquisition. But the inexperienced might be less easily confused about what the correct sample space was if they were aware of the general need for protocols for new information.

Insights in the foundations of the Bayesian theory. As I stressed in the introduction, a recognition of the need for protocols has important consequences for the de Finetti–Savage picture of subjective probability. Most importantly, it undermines Savage's claim that the Bayesian theory is normative.

In *The Foundations of Statistics* (1954), Savage tried to convince us that we should want our preferences to obey axioms that imply these preferences are ordered by expected values like those in games of chance. He tried to convince us even that these axioms express such basic canons of rationality that it is normative for us to make sure our preferences do satisfy them. But Savage seems to have taken the rule of conditioning for granted. In order to justify conditioning, we must add to Savage's axioms the assumption that we have a protocol for new information. And while we will want to have such a protocol, it hardly seems normative for us to have it. We may, by an act of will, adjust our preferences to make them obey Savage's rules, but it may take more than an act of will to find out in advance what the possibilities are for what we will learn.

The Bayesian claim to be normative must be abandoned. We must recognize that when we make a Bayesian probability calculation we are only constructing an argument. We are assessing the strength of our evidence in a particular problem by drawing an analogy with the evidence we would have for a particular outcome in a particular game of chance. This argument by analogy may be strong, but it is only an argument. One aspect of its strength is the extent to which we can claim there is a protocol in our problem like the protocol that would be present in the game of chance.

A recognition of the general need for protocols can also help clarify several other issues that have arisen in philosophical discussions of the Bayesian theory.

Consider, for example, the recent discussions of conditioning by Paul Weirich (1983) and Bas van Frassen (1983). These authors give examples where conditioning can lead to contradictions when it is applied to a person's beliefs about his own future beliefs or actions. Weirich uses such examples to argue that some events cannot represent all one has learned and hence cannot be candidates for conditioning. This important insight is bolstered by recognition of the need for a protocol, for it is clear that a given protocol can justify conditioning on only some of the events in a field of events. (See the discussion of exact events in Appendix 1.)

Consider also the debate over the 'likelihood principle'. According to this principle, statistical inferences should depend only on the element of the sample space actually observed. They should not depend on whole sample space. Though it was originally thought of as a consequence of other more intuitively meaningful principles, the likelihood principle has been presented in recent years as if it were self-evident; non-Bayesian methods which do not obey it have been criticized because they 'depend on what might have been observed'. The force of this rhetoric is diminished when we recognize that Bayesian conditioning, which does satisfy the likelihood principle, depends for its legitimacy on a protocol, i.e. on a specification of what might have been observed (Lindley, 1980).

Finally, consider the Bayesian criticism of Dempster's rule of combination in the theory of belief functions. Several authors, for example, Seidenfeld (1981), have claimed that the theory of belief functions is less flexible than the Bayesian theory because Dempster's rule is appropriate only for combining independent items of evidence, whereas Bayesian probabilities can be conditioned on arbitrary new evidence. But when we recognize the need for protocols in the Bayesian theory, we see that this theory has just as much need to model the evidence as the theory of belief functions does.

An insight in mathematical probability. Almost all students of mathematical probability are initially puzzled by the difference between Kolmogorov's treatment of conditional probability in the discrete case and his treatment of it in the general case. In the discrete case, we are allowed to condition on any event B such that $\Pr\{B\}>0$ using the definition (1). But in the general case, where it is sometimes necessary to condition on events of probability zero, we are told first that it is meaningful to consider an event B only when B is thought of as an element of a partition of the sample space, and secondly that the resulting conditional probabilities depend on the particular partition chosen. This puzzlement is partially relieved when it is explained that continuous probability distributions are only idealizations and must be discretized to correspond to reality; the partition serves to indicate what discretization is intended. The example of conditioning a distribution over a sphere on a particular longitude is often cited; in reality the longitude can only be measured with error, and so the result of the conditioning should depend on how the error varies with the latitude (Bertrand, 1889, pp. 6–7; Borel, 1909, pp. 100–104; Kolmogorov, 1933, pp. 44–45). But some puzzlement remains. If conditioning on an isolated event is meaningful in the discrete case, why is it not meaningful in the general case?

The answer, of course, is that conditioning on an isolated event is not meaningful even in the discrete case. Conditioning is legitimate and meaningful only when there is a protocol which specifies what else might have been observed and hence determines a partition (or at least a system of partitions; see Appendix 1).

Do frequentists need protocols? Kolmogorov's treatment of probability was inspired by a frequentist rather than a subjective conception of probability. Do protocols have a role to play in a frequentist theory of probability?

The answer seems to be positive if we agree that the frequentist view is based on the idea of random sampling from a population. If, after we learn that our selection is in a certain subset of the population, we want to think of ourselves as sampling at random from that subset, then we need to be assured that our learning that it is in the subset does not depend in any way on where in the subset it is.

Acknowledgments

The author has benefited from conversation and correspondence with many people, including Ben Cobb, Art Dempster, Michael Goldstein, Ian Hacking, David Krantz, Paul Meir, Saul Spatz, Terry Speed, Steve Stigler,

Consider also the debate over the 'likelihood principle'. According to this principle, statistical inferences should depend only on the element of the sample space actually observed. They should not depend on whole sample space. Though it was originally thought of as a consequence of other more intuitively meaningful principles, the likelihood principle has been presented in recent years as if it were self-evident; non-Bayesian methods which do not obey it have been criticized because they 'depend on what might have been observed'. The force of this rhetoric is diminished when we recognize that Bayesian conditioning, which does satisfy the likelihood principle, depends for its legitimacy on a protocol, i.e. on a specification of what might have been observed (Lindley, 1980).

Finally, consider the Bayesian criticism of Dempster's rule of combination in the theory of belief functions. Several authors, for example, Seidenfeld (1981), have claimed that the theory of belief functions is less flexible than the Bayesian theory because Dempster's rule is appropriate only for combining independent items of evidence, whereas Bayesian probabilities can be conditioned on arbitrary new evidence. But when we recognize the need for protocols in the Bayesian theory, we see that this theory has just as much need to model the evidence as the theory of belief functions does.

An insight in mathematical probability. Almost all students of mathematical probability are initially puzzled by the difference between Kolmogorov's treatment of conditional probability in the discrete case and his treatment of it in the general case. In the discrete case, we are allowed to condition on any event B such that $\Pr[B]>0$ using the definition (1). But in the general case, where it is sometimes necessary to condition on events of probability zero, we are told first that it is meaningful to consider an event B only when B is thought of as an element of a partition of the sample space, and secondly that the resulting conditional probabilities depend on the particular partition chosen. This puzzlement is partially relieved when it is explained that continuous probability distributions are only idealizations and must be discretized to correspond to reality; the partition serves to indicate what discretization is intended. The example of conditioning a distribution over a sphere on a particular longitude is often cited; in reality the longitude can only be measured with error, and so the result of the conditioning should depend on how the error varies with the latitude (Bertrand, 1889, pp. 6–7; Borel, 1909, pp. 100–104; Kolmogorov, 1933, pp. 44–45). But some puzzlement remains. If conditioning on an isolated event is meaningful in the discrete case, why is it not meaningful in the general case?

The answer, of course, is that conditioning on an isolated event is not meaningful even in the discrete case. Conditioning is legitimate and meaningful only when there is a protocol which specifies what else might have been observed and hence determines a partition (or at least a system of partitions; see Appendix 1).

Do frequentists need protocols? Kolmogorov's treatment of probability was inspired by a frequentist rather than a subjective conception of probability. Do protocols have a role to play in a frequentist theory of probability?

The answer seems to be positive if we agree that the frequentist view is based on the idea of random sampling from a population. If, after we learn that our selection is in a certain subset of the population, we want to think of ourselves as sampling at random from that subset, then we need to be assured that our learning that it is in the subset does not depend in any way on where in the subset it is.

Acknowledgments

The author has benefited from conversation and correspondence with many people, including Ben Cobb, Art Dempster, Michael Goldstein, Ian Hacking, David Krantz, Paul Meir, Saul Spatz, Terry Speed, Steve Stigler,

Peter Walley and Sandy Zabell. The editor and referees were also helpful. Appendix 1 has benefited a great deal from suggestions by Peter Williams.

The author's research was partially supported by a Guggenheim fellowship and by NSF grant MCS 830-1282.

Appendix 1. The mathematics of protocols

In § 3 we limited our attention to two events A and B and to a protocol for your learning about the happening of B. It is also enlightening to think about more complete protocols, protocols that specify the possibilities in a more extended learning process. This appendix is devoted to a mathematical study of such protocols.

Here, as in § 3, we are interested in a protocol for your subjective learning process, not a protocol that orders objective events that may or may not come to someone's attention. This protocol may be purely subjective, or it may be publicly agreed to and thus objective as well. In any case, an event in the protocol is always the event that you learn something.

We begin by developing the idea of a complete protocol in terms of a tree. We then give a more abstract treatment.

Trees. Let Ω denote the set of possibilities for how your knowledge will grow during a certain learning process. These possibilities can be represented graphically as different ways of moving down an upside-down tree, as in Fig. 1. You begin at o, the initial node of the tree, and move down step by step.

If the tree is finite, as in the figure, then each path down the tree can be identified by its terminal node. Thus we may think of Ω, the set of all paths down the tree, as the set of all terminal nodes.

An *event* is a subset of Ω. An event A *happens* when you take a step down the tree that forces your final path to be in A. In Fig. 1, for example, the event $B=\{b, c, d\}$ happens as you arrive at node n. Other events also happen at this point, e.g. $\{b, c, d, e, f, g\}$. But the event B is of particular interest because it is more specific than the other events that happen as you arrive at node n. It represents, as it were, everything that has happened (i.e. everything you have learned) at that point. It is the most specific event that happens as you arrive at node n because it consists of all the terminal nodes that lie below n. (Notice also that B can happen only as you arrive at node n. Other events can happen in several ways. The event $\{b, c, d, e, f, g\}$, for example, can happen as you arrive at node n or as you arrive at node p.) Let us call an event that consists of all the terminal nodes below a given node an *exact* event. It is 'exactly' what happens as you arrive at that node.

The concept of an exact event clarifies the assumptions about the event B that we made in § 3. For within the framework of a complete protocol, those assumptions reduce to the

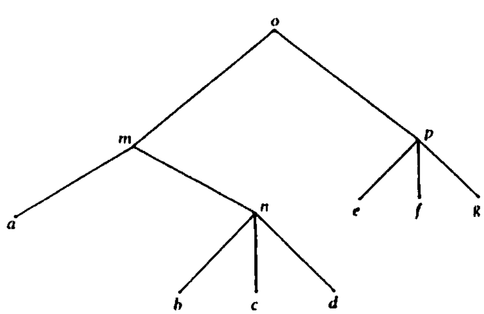

Figure 1. *A protocol represented by a tree.*

Peter Walley and Sandy Zabell. The editor and referees were also helpful. Appendix 1 has benefited a great deal from suggestions by Peter Williams.
The author's research was partially supported by a Guggenheim fellowship and by NSF grant MCS 830-1282.

Appendix 1. The mathematics of protocols

In § 3 we limited our attention to two events A and B and to a protocol for your learning about the happening of B. It is also enlightening to think about more complete protocols, protocols that specify the possibilities in a more extended learning process. This appendix is devoted to a mathematical study of such protocols.

Here, as in § 3, we are interested in a protocol for your subjective learning process, not a protocol that orders objective events that may or may not come to someone's attention. This protocol may be purely subjective, or it may be publicly agreed to and thus objective as well. In any case, an event in the protocol is always the event that you learn something.

We begin by developing the idea of a complete protocol in terms of a tree. We then give a more abstract treatment.

Trees. Let Ω denote the set of possibilities for how your knowledge will grow during a certain learning process. These possibilities can be represented graphically as different ways of moving down an upside-down tree, as in Fig. 1. You begin at o, the initial node of the tree, and move down step by step.

If the tree is finite, as in the figure, then each path down the tree can be identified by its terminal node. Thus we may think of Ω, the set of all paths down the tree, as the set of all terminal nodes.

An *event* is a subset of Ω. An event A *happens* when you take a step down the tree that forces your final path to be in A. In Fig. 1, for example, the event $B=\{b, c, d\}$ happens as you arrive at node n. Other events also happen at this point, e.g. $\{b, c, d, e, f, g\}$. But the event B is of particular interest because it is more specific than the other events that happen as you arrive at node n. It represents, as it were, everything that has happened (i.e. everything you have learned) at that point. It is the most specific event that happens as you arrive at node n because it consists of all the terminal nodes that lie below n. (Notice also that B can happen only as you arrive at node n. Other events can happen in several ways. The event $\{b, c, d, e, f, g\}$, for example, can happen as you arrive at node n or as you arrive at node p.) Let us call an event that consists of all the terminal nodes below a given node an *exact* event. It is 'exactly' what happens as you arrive at that node.

The concept of an exact event clarifies the assumptions about the event B that we made in § 3. For within the framework of a complete protocol, those assumptions reduce to the

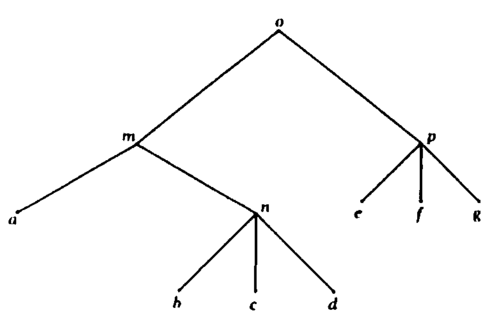

Figure 1. *A protocol represented by a tree.*

assumption that B is exact. An event is exact if you expect that its happening, if and when it happens, will represent all that you will have learned up to that point.

It is natural, in the context of a protocol defined by a tree, to identify your possible future probabilities by referring to nodes of the tree rather than by referring to events that have happened. In Fig. 1, for example, with $B=\{b, c, d\}$, we might use the symbol $\Pr_n[A]$, rather than $\Pr[A \mid B]$ to denote the number you expect your probability for A to be if and when you arrive at node n. If we similarly use the subscript o to denote your probabilities at the initial node o, then the argument of § 3 will yield the conclusion

$$\Pr_o[A \cap B] = \Pr_o[B] \Pr_n[A].$$

For details see Shafer (1982).

This argument justifies the rule of conditioning only in the case of exact events. But the question of conditioning on other events does not arise. You expect you will always find yourself at a node, so you expect you will always want to condition on an exact event.

Notice that the complement of an exact event need not be exact. (In Fig. 1, for example, $\{b, c, d\}$ is exact, but its complement $\{a, e, f, g\}$ is not.) Thus it is not necessary for the argument for conditioning on an event that there should be a single point in the future where you expect to learn of the happening or failure of the event; compare Skyrms (1980, p. 120.)

Axioms for expectations. The structure of a tree can be described merely by specifying the exact events. This suggests the following definition.

Definition. A *protocol* is a nonempty collection $\mathcal{S}$ of subsets of a nonempty set Ω such that

(i) any two elements of $\mathcal{S}$ are either disjoint or nested, and
(ii) if $\omega \in \Omega$, $S \in \mathcal{S}$, and $\omega \notin S$, then there is an element of $\mathcal{S}$ that contains ω and is disjoint from S.

The elements of $\mathcal{S}$ correspond to the exact events in a tree. Here let us call them *situations.*

Let us call a bounded real-valued function X on Ω an *expectation* for the protocol if for every ω in Ω there is a situation S containing ω such that X is constant on S. (This means that no matter what happens there will eventually be a situation where the value of X is settled.) The expectations for a protocol form an algebra, which we denote by $\mathcal{X}$. Let us call a subset of Ω an *event* if its indicator (the function that is one on the event and zero off the event) is an expectation. The events for a protocol form a field. Following de Finetti (1972, p. xviii), we may allow any symbol that designates an event to also designate its indicator.

An *evaluation* for a protocol is a real-valued function E on $\mathcal{X} \times \mathcal{S}$ that satisfies the following axioms for all S in $\mathcal{S}$ and all $X, X_1, \ldots, X_n$ in $\mathcal{X}$.

Axiom A. If $X(\omega)=c$ for all ω in S, then $E[X \mid S]=c$.

Axiom B. If for every element ω of S there is a situation S' such that $\omega \in S' \subseteq S$ and $\sum_i E[X_i \mid S'] \geq c$, then $\sum_i E[X_i \mid S] \geq c$, where the sums are over $i=1, \ldots, n$ (and similarly for $\leq$).

Axiom B is based on the idea that $E[X \mid S]$ is a fair price for X in the situation S. Fair prices should not allow a scheme of buying and selling that guarantees a definite profit. If, for example, $E[X \mid S] < c$ and yet for every element $\omega \in S$ there is a situation S' such that $\omega \in S' \subseteq S$ and $E[X \mid S'] \geq c$, then a person in situation S can assure himself of the definite profit $c - E[X \mid S]$ by buying the expectation X at the price $E[X \mid S]$ and reselling it when the price reaches c.

Axioms A and B imply the following axiom, which is familiar from the work of de Finetti.

Axiom C. If $\sum_i X_i(\omega) \geq c$ for all $\omega \in S$, then $\sum_i E[X_i \mid S] \geq c$, where the sums are over $i = 1, \ldots, n$ (and similarly for $\leq$).

As de Finetti has shown (de Finetti, 1974, pp. 74–75), Axiom C implies the usual properties for $E[.\mid S]$:

$$\inf_{\omega \in S} X(\omega) \leq E[X \mid S] \leq \sup_{\omega \in S} X(\omega),$$
$$E[aX \mid S] = aE[X \mid S], \quad E[X + Y \mid S] = E[X \mid S] + E[Y \mid S]. \tag{A1}$$

In addition to these properties, Axiom B also implies that

$$E[S_2 X \mid S_1] = E[S_2 \mid S_1] E[X \mid S_2] \tag{A2}$$

whenever $S_2 \subset S_1$.

If A is an event, then the quantity $E[A \mid S]$ may be called the *probability* of A in situation S, and it may also be denoted by $\Pr[A \mid S]$. The properties given in (A1) for $E[.\mid S]$ yield the usual rules for $\Pr[.\mid S]$:

$$\Pr[\phi \mid S] = 0, \quad \Pr[S \mid S] = 1, \quad \Pr[A \cup B \mid S] = \Pr[A \mid S] + \Pr[B \mid S] - \Pr[A \cap B \mid S].$$

And (A2) yields

$$\Pr[A \cap S_2 \mid S_1] = \Pr[S_2 \mid S_1] \Pr[A \mid S_2],$$

the rule of compound probability.

Suppose π is a partition of Ω into situations; i.e., a subset of $\mathscr{S}$ whose elements are pairwise disjoint and have Ω as their union. For each ω in Ω, let us denote by $S_\pi(\omega)$ the element of π that contains ω. And for each expectation X, let us denote by $E_\pi(X)$ the expectation whose value at ω is $E[X \mid S_\pi(\omega)]$. We see from (A1) that the operator $E_\pi(.)$ is linear:

$$E_\pi(aX) = aE_\pi(X), \tag{A3}$$

$$E_\pi(X + Y) = E_\pi(X) + E_\pi(Y). \tag{A4}$$

From (A3) and Axiom C we see that

$$E_\pi(XY) = XE_\pi(Y) \tag{A5}$$

whenever X is constant on each situation in π. And we see from Axiom B directly that

$$E_{\pi_1}(E_{\pi_2}(X)) = E_{\pi_1}(X) \tag{A6}$$

whenever π_2 is a refinement of π_1. This is a form of the *rule of iterated expectation*. For related derivations of this rule, see Goldstein (1983) and Shafer (1983).

A law of large numbers. As an illustration of the potential of the framework we have just laid out, let us prove a version of the law of large numbers.

Suppose a person intends to buy successively a sequence $X_1, X_2, \ldots$ of uniformly bounded expectations, selling each when he buys the next. He has definite intentions about when he will buy each expectation. For simplicity, suppose Ω is a situation, the person is initially in Ω, and he buys X_1 while still in Ω. Then we may spell out his further intentions by specifying a sequence $\pi_1, \pi_2, \ldots$ of successively finer partitions of Ω into situations; in the situations in π_i he will sell X_i and buy X_{i+1}.

What can we say about the person's average gain from the first n expectations? His gain from X_i is

$$G_i = E_{\pi_i}(X_i) - E_{\pi_{i-1}}(X_i);$$

here $\pi_0 = \{\Omega\}$. Using (A4), (A5) and (A6), we find that $E[G_i \mid \Omega] = 0$ and $E[G_i G_j \mid \Omega] = 0$ for $i \neq j$. So the usual proof yields the law of large numbers for the G_i: for every $\delta > 0$ and $\varepsilon > 0$, there exists N such that

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} G_i\right| \leq \delta \mid \Omega\right] \geq 1 - \varepsilon \tag{A7}$$

for all $n > N$.

We obtain an interesting special case of (A7) if we take the expectation X_i to be an event, say A_i, and if we assume that A_i is always determined before the expectation is sold. In this case,

$$G_i = A_i - \Pr_{\pi_{i-1}}[A_i],$$

$$\frac{1}{n}\sum_{i=1}^{n} G_i = \frac{1}{n}\sum_{i=1}^{n} A_i - \frac{1}{n}\sum \Pr_{\pi_{i-1}}[A_i].$$

So (A7) tells us that if a person intends to announce his subjective probabilities for a long sequence of events, always waiting for one event to be determined before he gives his probability for the next, then he expects $\sum A_i/n$, the proportion of the events that happen, to be close to $\sum \Pr_{\pi_{i-1}}[A_i]/n$, the average of the probabilities he announces. This is Dawid's calibration theorem (Dawid, 1982).

Conglomerability. Axiom B is related to the property of *conglomerability*, introduced by de Finetti (1972, p. 177–178) and studied by Dubins (1975) and others. There are several different ways of making the concept of conglomerability precise, but roughly speaking, Axiom B says that the probability distribution $\Pr[.\mid S]$ must be conglomerative with respect to every partition of S into situations.

Discussion of conglomerability has centered around the fact that a probability distribution cannot be expected to be conglomerative with respect to all partitions. If it is not countably additive, it cannot even be expected to be conglomerative with respect to all countable partitions. De Finetti has argued that the intuition that supports conglomerability must therefore be given up. But once a protocol is brought into the picture, we see that this intuition applies only to partitions into situations, not to all partitions. Axiom B demands conglomerability only for partitions into situations.

Appendix 2. De Finetti's argument for the rule of compound probability

Since the 1930's Bruno de Finetti has expounded an approach to conditional probability that dispenses with the idea of a protocol but nonetheless treats the rule of compound probability as a theorem.

Our argument in § 3 above was based on a protocol that says that if B happens you will learn of its happening and nothing else. In the context of this protocol, we were able to derive the rule of compound probability with $\Pr[A \mid B]$ defined as the probability you expect to have for A if and when you learn of B's happening. De Finetti takes a different tack. He defines $\Pr[A \mid B]$ as the price at which you will buy or sell a contract that pays \$1 if A is true, with the understanding that the purchase will be cancelled unless B is true.

This definition allows him to reason as follows.

> By paying $\$\Pr[A \cap B]$, you can by a contract that returns \$1 if $A \cap B$ is true. But you can obtain the same result by arranging to pay $\$\Pr[A \mid B]$ in case B is true for a contract that returns \$1 if A is also true, and you can arrange to have this $\$\Pr[A \mid B]$ if it is needed by paying now the amount $\$\Pr[B]\Pr[A \mid B]$ for a contract that returns $\$\Pr[A \mid B]$ if B is true. Hence $\Pr[A \cap B] = \Pr[B]\Pr[A \mid B]$.

This is a paraphrase of an argument given on p. 135 of de Finetti (1974). For more formal proofs of the rule of compound probability from de Finetti's viewpoint, see de Finetti (1974, p. 136–139), de Finetti (1972, p. 15–16), or de Finetti (1980, p. 69).

One way to see that there is a difference between de Finetti's argument and the argument based on a protocol is to note that de Finetti's argument applies to any pair of events A and B. De Finetti repeatedly insists on this point; see, for example, de Finetti (1972, p. 193). The argument based on a protocol is not so general. It applies only to exact events, and as we saw in Appendix 1, not every event is exact for a given protocol.

By dispensing with protocols, de Finetti loses the logic that ties the argument to your expectations about your future probabilities. If we consent to de Finetti's use of the symbol '$\Pr[A \mid B]$' to denote the price of a contract that returns \$1 if A is true but whose purchase is cancelled if B is false, then we can agree with de Finetti that $\Pr[A \mid B] = \Pr[A \cap B]/\Pr[B]$. But this tells us nothing about what you expect your probability for A to be if you learn B. It is true that learning B will make the simple contract that pays \$1 if A is true equivalent for you to the conditional contract that pays \$1 if A is true but is cancelled unless B is true. But this only tells us that the prices of the two contracts should be the same after you have learned B is true. It does not tell us that you should expect their new common price to be $\Pr[A \cap B]/\Pr[B]$, the price you earlier assigned to the conditional contract. To pretend that the price for this conditional contract should be expected not to change, although the prices for other contracts do change, is to repeat the fallacy in Bayes's argument for his fifth proposition. For further discussion of this point, see Shafer (1981, pp. 33–38; 1982, pp. 1078–1080).

Because he refuses to bring the idea of protocols into the theory of subjective probability, de Finetti's derivation of the rule of compound probability takes us no farther than Kolmogorov's definition. It leaves the problem of judging when conditioning is legitimate still entirely outside the theory of subjective probability.

References

Ball, W.W.R. (1911). *Mathematical Recreations and Essays*, 5th ed. London: Macmillan.
Bar-Hillel, M. & Falk, R. (1982). Some teasers concerning conditional probability. *Cognition* **11**, 109–122.
Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. London for 1763* **53**, 370–418.
Bernoulli, J. (1713). *Ars Conjectandi*. Basel.
Bertrand, J. (1889). *Calcul des Probabilities*. Paris: Gauthier-Villars et Fils.
Borel, E. (1909). *Éléments de la Théorie des Probabilités*. Paris: A. Hermann & Fils.
Copi, I.M. (1968). *Introduction to Logic*, 3rd ed. New York: Macmillan.
Dale, A.I. (1974). Discussion: On a problem in conditional probability. *Philos. Sci.* **41**, 204–206.
Dawid, A.P. (1982). The well-calibrated Bayesian. *J. Am. Statist. Assoc.* **77**, 605–613.
Dawid, A.P. & Dickey, J.M. (1977). Likelihood and Bayesian inference from selectively reported data. *J. Am. Statist. Assoc.* **72**, 845–850.
de Finetti, B. (1972). *Probability, Induction and Statistics*. New York: Wiley.
de Finetti, B. (1974). *Theory of Probability*, **1**, New York: Wiley.
de Finetti, B. (1975). *Theory of Probability*, **2**. New York: Wiley.
de Finetti, B. (1980). Foresight: Its logical laws, its subjective sources. In *Studies in Subjective Probability*, 2nd ed., Eds. H.E. Kyburg, Jr. and H.E. Smokler, pp. 51–131. New York: Krieger.

De Moivre, A. (1738). *The Doctrine of Chances*, 2nd ed. London: A. Millar.
Dubins, L.E. (1975). Finitely additive conditional probabilities, conglomerability, and disintegrations. *Ann. Prob.* **3**, 89–99.
Faber, R. (1976). Discussion: re-encountering a counter-intuitive probability. *Philos. Sci.* **43**, 283–285.
Feller, W. (1950). *An Introduction to Probability Theory and its Applications*, **1**. New York: Wiley.
Freund, J.E. (1965). Puzzle or paradox? *Am. Statistician* **19** (4), 29–44. Related letters to the editor appeared in *Am. Statistician* **20**: (1), 34–36; (2), 37; (5), 36–37; and in **21**: (2), 42–44; (3), 38–41; (4), 51.
Gamow, G. & Stern, M. (1958). *Puzzle-Math*. New York: Viking.
Gardner, M. (1959). *The Scientific American Book of Mathematical Puzzles and Diversions*. New York: Simon and Schuster.
Goldberg, S. (1976). Discussion: Copi's conditional probability problem. *Philos. Sci.* **43**, 286–289.
Goldstein, M. (1983). The prevision of a prevision. *J. Am. Statist. Assoc.* **78**, 817–819.
Gridgeman, N.T. (1963). The pit of paradox. *The New Scientist* **20**, 462–465.
Huygens, C. (1660). *Van Rekeningh in Spelen van Geluck*. Amsterdam.
Kac, M. & Rubinow, S. I. (1977). Probability of the Pittsburgh deaths. *Science* **196**, 480.
Keynes, J.M. (1921). *A Treatise on Probability*. London: Macmillan.
Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.
Lindley, D. V. (1980). Discussion of a paper by B.M. Hill. In *Bayesian Statistics; Proceedings of the First International Meeting Held in Valencia (Spain), May 28 to June 2, 1979*, Eds. J.M. Bernado, M.H. De Groot, D.V. Lindley and A.F.M. Smith, p. 59. Valencia University Press.
Mosteller, F. (1965). *Fifty Challenging Problems in Probability with Solutions*. Reading, Massachusetts: Addison-Wesley.
Neustadt, R.E. & Fineberg, H.V. (1982). *The Epidemic That Never Was*. New York: Random House.
Rose, L.E. (1972). Countering a counter-intuitive probability. *Philos. Sci.* **39**, 523–524.
Rubens, J. (1976). Restricted Choice. In *The Official Encyclopedia of Bridge*, 3rd ed., Ed. R.L. Frey, pp. 376–381. New York: Crown.
Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
Rubin, D.B. (1978). Bayesian inference for causal effects: the role of randomization. *Ann. Statist.* **6**, 34–58.
Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley.
Schrödinger, E. (1947). The foundation of the theory of probability-I. *Proc. R. Irish Acad.* A **51**, 51–66.
Seidenfeld, T. (1981). Statistical evidence and belief functions. In *PSA 1978*, Vol. 2, Eds. P.D. Asquith and I. Hacking, pp. 478–489. East Lansing, Michigan: Philosophy of Science Association.
Shafer, G. (1981). Constructive probability. *Synthese* **48**, 1–60.
Shafer, G. (1982). Bayes's two arguments for the rule of conditioning. *Ann. Statist.* **10**, 1075–1089.
Shafer, G. (1983). A subjective interpretation of conditional probability. *J. Philos. Logic* **12**, 453–466.
Shafer, G. (1985). Belief functions and possibility measures. In *The Analysis of Fuzzy Information*, **1**, Ed. J.C. Bezdek. CRC Press.
Shafer, G. & Tversky, A. (1985). Languages and designs for probability judgement. *Cognitive Sci.* **9**, 309–339.
Skyrms, B. (1980). Higher order degrees of belief. In *Prospects for Pragmatism*, Ed. D.H. Mellor, pp. 109–137. Cambridge University Press.
Teller, P. (1973). Conditionalization and observation. *Synthese* **26**, 218–258.
Tukey, J.W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science* **198**, 679–684.
van Frassen, B.C. (1983). Shafer on conditional probability. *J. Philos. Logic.* **12**, 467–470.
Weirich, P. (1983). Conditional probabilities and probabilities given knowledge of a condition. *Philos. Sci.* **50**, 82–95.

Résumé

Les valeurs des espérances dans un jeu de hasard se transforment avec le déroulement du jeu. Ce déroulement est governé par un protocole, qui précise ce qui peut se passer à chaque pas. Ici on étudie l'idée d'un protocole des points de vue intuitifs, historiques, et mathématiques. Les protocoles sont d'importance dans la statistique et dans l'évaluation des probabilités subjectives parce que la signification d'un renseignement dépend des conditions qui gouvernent sa transmission. En présence d'un protocole, le conditionnement des probabilités est obligatoire; sans protocol, ce conditionnement est douteux.

[*Received November* 1983. *Revised December* 1984]

Discussion of paper by G. Shafer

T.P. Speed

CSIRO Division of Mathematics and Statistics, G.P.O. Box 1965, Canberra, ACT 2601, Australia

Glenn Shafer has presented us a compelling case for the reintroduction, formally or informally, of protocols into elementary probability, so that our understanding of what we are doing can return to the high level it attained in the mid-18th century. I can only applaud his paper: it clarifies some well-known puzzles, it points out inadequacies in the claim of Bayesians that their approach is normative without protocols, it offers insight into Kolmogorov's definition of conditional probability given a partition or σ-field, and it connects nicely with existing work on selectively reported or missing data problems in statistics.

For fans of puzzles such as those mentioned in § 2, I offer one more, the Monte (or Monty) Hall Puzzle (Selvin, 1975a,b), which is similar to the three prisoner puzzle, but with an extra twist.

I was disappointed not to see Glenn Shafer express the rule of compound probability (2) as

$$\Pr(A \,\&\, B \mid C) = \Pr(A \mid C) \Pr(B \mid A \,\&\, C).$$

In teaching I try to use this form to avoid problems resulting from the fact pointed out in the paper, that the sample space, relative to which probabilities are unconditional, plays only an implicit role in most discussions. Surely it is implicit features of the framework with which most care is needed, and which should be avoided if at all possible. Is this not Shafer's message?

A question which needs attention is the following: to what extent should teachers of elementary probability incorporate protocols into their syllabuses? I would be as reluctant to see the formal apparatus of protocols with trees, exact events, situations, etc, put into a first course, interesting though it is to the specialist, as I would to see the whole issue ignored.

Let us hope that some suitable way can be found whereby the wisdom of de Moivre and Bayes can, after more than two centuries, again be passed on to students of probability.

Additional references

Selvin, S. (1975a). A problem in probability. In Letters to the Editor. *Am. Statistician* **29**, 67.
Selvin, S. (1975b). On the Monty Hall problem. *Am. Statistician* **29**, 134.

[*Received February* 1985]

Reply to Discussion

Glenn Shafer

It is a pleasure to respond to Terry Speed's generous comments.

I heartily agree with Speed that we should not inflict protocols on beginning students.

Nor should we inflict Kolmogorov's axioms on them. A student's first encounter with probability should be through problems (Freudenthal, 1970).

In Appendix I of Shafer (1982), I repeated John Maynard Keynes's statement that Hugh McColl was the first to use an explicit notation for the probability of one proposition relative to another. In private correspondence, Terry Speed has pointed out to me that C.S. Peirce preceded McColl in this respect. In Peirce (1867) we find b_a used to denote the probability of b relative to a. Moreover, in Peirce (1878) we find a clear verbal statement of the rule that Speed has repeated in modern notation: $\Pr(A \,\&\, B \mid C) = \Pr(A \mid C) \Pr(B \mid A \,\&\, C)$.

I do not find symbols like b_a or $\Pr(A \mid C)$ as attractive as Speed does. Yes, we do need to emphasize that the probability of A depends on the protocol or sampling framework. Yes, the symbol $\Pr(A \mid C)$ can be used to make this point; we explain that C represents the sample space and hence, indirectly, all our assumptions about how the sampling is done. But after the student forgets this explanation, the symbol $\Pr(A \mid C)$ will seem to put A and C on equal footing; both are just sets, or events, or propositions. And then it is easy to suppose, wrongly, that $\Pr(A \mid C)$ means something for any such pair.

Additional references

Freudenthal, H. (1970). The aims of teaching probability. In *The Teaching of Probability and Statistics*, Ed. L. Råde, pp. 151-167. New York: Wiley Interscience.
Peirce, C.S. (1867). On an improvement in Boole's calculus of logic. *Proc. Am. Acad. Arts Sci.* **7**, 250-261.
Peirce, C.S. (1878). Illustrations of the logic of science. Fourth paper—The probability of induction. *Popular Sci. Monthly* **12**, 705-718.

[*Received May* 1985]