also important differences. It is difficult to use the verb "measure" without pretending that there is a well-defined property to be measured. Talk about canonical examples encourages a more constructive attitude.

One aspect of the constructive nature of Bayesian probability judgment, emphasized by Shafer and Tversky (1985), is the fact that we must construct our starting point. We must construct a probability distribution before we can condition it or multiply it by likelihoods. Bayesian theorists often assert categorically that every new experience must be treated in terms of its likelihood. Lindley, for example, declares that "an AI system faced with uncertainty about A_2 and experiencing A_1 has to update its uncertainty by considering how probable what it has experienced is, both on the supposition that A_2 is true, and that A_2 is false." But since a person may get around to constructing "initial" probabilities only after experiencing A_1, he or she has the option of treating A_1 as part of the evidence for those initial probabilities. Consider Lindley's investigator, who has discovered evidence that a criminal is left-handed. Instead of treating this evidence in terms of its likelihood, the investigator uses it directly in constructing a probability distribution.

There are problems, of course, where the construction can all be done in advance and then applied to many cases. GLADYS deals with this kind of problem; the same framework is applied to one patient after another. If I understand Spiegelhalter correctly, he believes that the bounded nature of expert systems means that this is the only kind of problem with which they can deal.

A finite system that permits construction can, however, deal with an unbounded range of situations. This is one of the fundamental points of the generative theory of grammar. The constructive nature of human reasoning makes us capable of exploring ever new realms of experience, and the ambition of AI is to duplicate this capability. Rule-based expert systems are one attempt to do so. These systems do not handle probabilistic reasoning very well, and many AI theorists would conclude from this that probabilistic reasoning has little role in genuine intelligence. In order to prove them wrong, we must do more than retreat to bounded systems like GLADYS. We must take the problem of automating construction seriously.

ADDITIONAL REFERENCE

SHAFER, G. (1986b). Savage revisited (with discussion). *Statist. Sci.* **1** 463–501.

Comment: A Tale of Two Wells

Dennis V. Lindley

The main issue is whether uncertainty should be described by probability, belief functions, or fuzzy logic; not just in artificial intelligence and expert systems, but generally. Are we to be probabilists, believers, or fuzzifiers? Or do we need some mixture of all three disciplines? To me, the important distinction between the methods rests on the rules of combination of uncertainty statements. Do we operate with the calculus of probability, the rules of belief functions, or with those of fuzzy logic? In my paper the challenge was made "that anything that can be done by these methods (belief functions and fuzzy logic) can better be done with probability." This reply will address one such challenge and I hope to show that Dempster's rule for belief functions does not behave as well as Bayes rule. My discussion is therefore chiefly addressed to Shafer and Zadeh. The omission of any discussion of Spiegelhalter's contribution arises because I agree substantially with it, and highly regard it. I wish that his program for dyspepsia had been more Bayesian and that he had recognized that uncertainty about a probability is usually a reference to the desirability of obtaining more data, so that his conflict ratio should really reflect this. To return to the challenge.

In 1685 the then Bishop of Bath and Wells wrote a paper in which the following problem was discussed. Two witnesses separately report that an event is true. Both are known to be unreliable to the extent that they only tell the truth with probabilities p_1 and p_2 respectively. What reliability can we then place, in the light of the witnesses' testimonies on the truth of the event? The Bishop's answer was $1 - (1 - p_1)(1 - p_2)$. The following is a precis of his argument. If the event is false, both witnesses must have lied, an event of probability $(1 - p_1)(1 - p_2)$. Consequently one minus this is the required reliability.

The result retains its interest today because the

Bishop's rule of combination of the two pieces of evidence is the same as Dempster's rule used in belief-function calculus. So here we have a challenge: I maintain probability can do better. (Readers will notice that the example is similar to that of Slippery Fred, used by Shafer in his paper, but is somewhat simpler. It was introduced by Shafer in the oral discussion of the original papers.)

Almost 80 years later, in 1763, the rector of Tunbridge Wells, Thomas Bayes, introduced his rule, presumably being unaware of the Bishop's proposal. This is now known as Bayes' rule (of probability), which we now apply to the Bishop's problem.

Let A denote the event whose truth is in question, and write a_1 and a_2 for the statements by the two witnesses that A is true. Since A is uncertain and a_1, a_2 are known assertions, we have to calculate $p(A \mid a_1, a_2)$, the probability that A is true, given both a_1 and a_2, using the rules of the probability calculus. This probability, and all those subsequently calculated, are judgments by some person. When, later, the Bishop's values, p_1 and p_2, are used, it will be supposed that, suitably interpreted, they are accepted as his by this person.

It is easier to work with the odds rather than the probability. These satisfy Bayes' rule

$$(1) \qquad \frac{p(A \mid a_1, a_2)}{p(\bar{A} \mid a_1, a_2)} = \frac{p(a_1, a_2 \mid A)p(A)}{p(a_1, a_2 \mid \bar{A})p(\bar{A})}.$$

On the far right we have the original odds on A before the witnesses gave their evidence. Write $p(A) = \pi$, so that the odds are $\pi/(1 - \pi)$. Also on the righthand side, in the numerator, occurs the probability $p(a_1, a_2 \mid A)$. This is the probability, were A true, that *both* witnesses would report it so; that is, tell the truth. The problem as formulated tells us nothing about this but there is a strong hint of independence in the original presentation—notice the multiplication $(1 - p_1)(1 - p_2)$—so if it is presumed here we might write $p(a_1, a_2 \mid A) = p(a_1 \mid A)p(a_2 \mid A)$, and similarly in the denominator, $p(a_1, a_2 \mid \bar{A}) = p(a_1 \mid \bar{A})p(a_2 \mid \bar{A})$.

Next consider one term in the numerator, $p(a_1 \mid A)$. This is the probability that the first witness will say the event is true when indeed it is true: in other words, tell the truth. But this is not the only way he could tell the truth: he could announce A was false when indeed it was false. This is $p(\bar{a}_1 \mid \bar{A}) = 1 - p(a_1 \mid \bar{A})$, which occurs in the denominator. The Bishop's argument used p_1, the probability of telling the truth, and to apply the rector's approach it is necessary to relate p_1 to $p(a_1 \mid A)$ and $p(\bar{a}_1 \mid \bar{A})$. If t_1 is the event that the first witness tells the truth, then

$$p(t_1) = p(t_1 \mid A)p(A) + p(t_1 \mid \bar{A})p(\bar{A}).$$

But t_1 when A is true (false) is $a_1(\bar{a}_1)$, so

$$p_1 = p(a_1 \mid A)\pi + p(\bar{a}_1 \mid \bar{A})(1 - \pi)$$

on inserting the Bishop's value p_1 for $p(t_1)$. The simplest assumption is that $p(a_1 \mid A) = p(\bar{a}_1 \mid \bar{A})$; that is, truth is just as likely when A is true as when it is false. It then easily follows that the common value is p_1.

Applying the same argument to the second witness, we easily have from (1) that

$$\frac{p(A \mid a_1, a_2)}{p(\bar{A} \mid a_1, a_2)} = \frac{p_1 p_2 \pi}{(1 - p_1)(1 - p_2)(1 - \pi)},$$

whence,

$$(2) \qquad p(A \mid a_1, a_2) = \frac{p_1 p_2 \pi}{p_1 p_2 \pi + (1 - p_1)(1 - p_2)(1 - \pi)}.$$

It is this result that can be compared to the Bishop's $1 - (1 - p_1)(1 - p_2)$.

To reach the Bayesian result (2) some assumptions have been made. We list these and comment upon them.

I: $p(A) = \pi$ is known, and relevant to the answer.

Its relevance seems indisputable. Even the testimony of very reliable witnesses (p_1 and p_2 near 1) would leave some doubt in a person's mind about A if initially he thought it most improbable (small π). Conversely, unreliable witnesses would still leave him having appreciable probability for A if initially π was near 1. Since it is relevant, its value must be included in the calculations. This is perhaps the Bishop's main mistake: to fail to appreciate the importance of π.

II: a_1 and a_2 are independent, both given A, and given $\bar{A}$.

Notice that the independence assumption is quite subtle. It demands independence both when the event is true and when it is false—but not unconditionally. It is easy to imagine circumstances where one independence holds but not the other. Suppose A is the event that a defendant in a court of law truly committed the crime with which he has been charged. If A is true, two witnesses may collude in providing him with an alibi; if A is false, no such collusion is needed. So a_1 and a_2 may be independent given $\bar{A}$, but not given A.

The Bishop almost certainly was tacitly assuming independence in 1685. It is also supposed in the modern belief function treatment, and Dempster's rule only realistically applies when it obtains. The

Bayesian approach works without independence: it has only been assumed here for simplicity and comparison with beliefs. What the Bayesian view does is to force one to consider the subtle nature of the dependence between the witnesses.

III: $p(a_i \mid A) = p(\bar{a}_i \mid \bar{A}), \quad (i = 1, 2).$

This asserts that the witnesses are equally reliable whether A is true or false. Again it is easy to imagine circumstances where this is not true. In some cultures there is a tendency for witnesses to say what they think will please the listener. So if A is the event "the airport is near," veracity is more likely when A is true than when it is false. Consequently one cannot be sure that $p(a_i \mid A)$ and $p(\bar{a}_i \mid \bar{A})$ are both p_i.

The Bishop certainly did not recognize the distinction, as have many writers after him. The Bayesian approach does not demand the equality: it merely forces one to recognize that two types of veracity are possible.

Applied to the Bishop's problem, the rector's approach forces one to consider one's initial belief in the event, the nature of the dependence between the witnesses, and the two forms of reliability that arise. We suggest that, on reflection, it will be admitted that all three features are relevant to the final answer. Even if the independencies and the equalities of the reliabilities are admitted, as the Bishop and the modern equivalent tacitly do, the result is still different from the Bishop's. It is of interest to enquire when they are equal. Equating (2) and $1 - (1 - p_1)(1 - p_2)$ easily gives after a little algebra the condition that

$$(1 - \pi) = p_1 p_2 \pi + (1 - p_1)(1 - p_2)(1 - \pi).$$

The righthand side is $p(a_1, a_2)$, the unconditional probability that both witnesses assert A is true, so that the Bishop and rector only agree (under assumptions II and III) if

$$p(\bar{A}) = p(a_1, a_2).$$

In words, the probability that the event is false has to be equal to the probability that both witnesses assert its truth. This is surely unreasonable.

I put it to the readership: my challenge has survived, probability does do better. Let us support the rector of Tunbridge Wells and not the Bishop of Bath and Wells: let us favor truth and not the establishment. (Bayes was a minister in the unestablished church.)

ACKNOWLEDGMENTS

I am grateful to Richard E. Barlow for useful comments on a first version of this tale, to Sir Richard Eggleston for illuminating discussions on the legal problems with two witnesses, and to Glenn Shafer for drawing my attention to the Bishop's paper.

Comment

David J. Spiegelhalter

It is fairly predictable that I should agree wholeheartedly with Professor Lindley's lucid justification of probability as the correct paradigm for handling uncertainty in expert systems (but how strange it is to see him cast in the role of defender of orthodoxy!). In particular, his emphasis on remembering the background evidence H is crucial to avoid any conception that there is a single "true" probability of an event, and the frequent references to the operational meaning of probability gives a practical as well as a theoretical justification. However, playing the devil's advocate, I see two main reasons why the artificial intelligence community may not be convinced by the argument.

Firstly, he turns all statements expressing uncertainty into expressions of probability concerning (at least theoretically) verifiable events, whereas many constructors of expert systems would prefer to keep their propositions deliberately imprecisely defined in order to look more like human reasoning, and do not provide an operational means of verification. Secondly, even if verifiable events *are* being considered, the scoring rule argument presumes a certain type of evaluation procedure which many might claim was rarely appropriate, since the criteria for the "success" of an expert system may only require a very coarse handling of uncertainty.

Nevertheless, the theoretical arguments concerning optimality and coherence are only one weapon in the armoury. Pearl (1986b), in a recent strong advocacy of probability, uses no normative criteria but concentrates on the power of the theory in adequately modeling complex evidential reasoning, and I feel, in the end, it will be the intuitive appeal and flexibility of probabilistic reasoning that will change the current climate.