Comment

D. V. Lindley

1. STATISTICS

Maximization of expected utility (MEU) has so many important implications for statistics that an examination of one of its axiomatic foundations by as careful and original a scholar as Shafer is to be welcomed. He is critical of the axioms and I fear that many statisticians sensing this will draw the conclusion that Shafer has undermined the Savage axioms and that therefore MEU, the likelihood principle, and Bayesian statistics can be forgotten. They need an excuse to forget and get on with their unbiased estimates, tail-area significance tests, and confidence limits. It is therefore important to notice that Shafer's penetrating criticisms are not carried through to produce an alternative axiomatization, despite the hints to this effect at the beginning of the paper. We may hazard a guess that he feels that belief functions provide a possible substitute for MEU, but these, or any other system known to me, do not imply that currently popular methods of statistical analysis are sound. They are silent, for example, on the basic issue of the likelihood principle. In fact, he suggests that, where MEU is sensibly based on analogies with games of chance, it is sound and therefore the principle applies. So Bayesian statistics survives.

A second point to be recorded before passing to the central issue I wish to discuss, is my complete disagreement with Shafer's third paragraph. It was not until the late 1950s that Savage appreciated the Bayesian implications of what he had done: prior to that he had looked upon MEU as a foundation for sampling-theory statistics. Surely it is wrong to say that "the need for subjective judgment is now widely understood." Very few papers in statistical journals incorporate subjective views, although the number is increasing. Again it is wrong to say that MEU is obstructive; it is very constructive. Workers in artificial intelligence and expert systems are beginning to realize that an intelligent expert ought to think probabilistically.

2. PSYCHOLOGICAL EXPERIMENTS AND PARADOXES

Shafer makes much of the work of psychologists who have carried out experiments showing that people do not maximize expected utility. It should be remembered that in almost all these experiments the subjects are students, required to assess probabilities when they have had no instruction in probability, or required to make decisions in trivial situations that are of no real importance to them. Is it really surprising that they are not very good at probability assessment or decision making? I draw quite a different conclusion from Shafer's. The bad nature of the inferences made and actions taken suggests that MEU has an enhanced status; for were it to be adopted, then there might well be a substantial improvement in decision making in fields where it really matters—and we all know that an improvement is needed. Had the psychologists' subjects been good maximizers the normative theory would have had little to offer.

Shafer also emphasizes the role of the paradoxes in MEU. He fails to point out that MEU can accommodate certain types of paradoxical behavior. Let us take Raiffa's (Figure 1) brilliant critique of Allais' paradox (Table 7). The only difference between Allais' original choice between f' and g' (at the left-hand edge of the tree in Figure 1) and Raiffa's suggested choice (after the white ball has been drawn) is, of course, the drawing of the white ball, the possible disappointment that it was not orange, and that \$500,000 has passed one by. If the utility for Raiffa's choice reflects this disappointment then when we turn to f and g (where the underlined \$500,000 is replaced by zero) no such disappointment is felt and the judgment may be different. I suspect that it often happens that when a person's behavior appears paradoxical it is because he is taking into account something that you have not considered and he has not mentioned. (In this example, the disappointment.) Readers may like to consider whether such an effect is really relevant in Allais' case. I think it is not.

3. NORMATIVE IDEAS

The relationship between normative and empirical concepts is a subtle one. I would like to argue by historical analogy. It is an analogy that I have used repeatedly before but it seems useful to me, and the critiques of it have not substantially changed its relevance for me to MEU. We have a normative theory for distances on the Earth's surface called (three-dimensional) geometry. This is basically due to Euclid. For many centuries this was little used because of the difficulties of measuring distances. Consider, for example, the great error that supreme navigator Colum-

D. V. Lindley was, until his retirement in 1977, Head of the Department of Statistics at University College, London. His mailing address is 2 Periton Lane, Minehead, Somerset TA24 8AQ, England.

bus made in the determination of longitude, leading him to confuse America and Asia. It was not until good methods of measurement allied to a sound method of handling them—triangulation—that Euclidean methods were successful. Even today there are discrepancies between the theory and actual measurements so that even one of the world's best triangulations reveals a slight mismatch and the distance between the extremities of the British Island may be out by a few millimetres.

Surely we should not demand more of the apparently much more difficult task of measuring peoples' beliefs and values than we do of distances on the Earth's surface. All of us who have walked in wild country know how misleading distance observed by eye can be, and the great value of a good map. With subjective probability, we are today only in the position corresponding to measurement by eye: we have no maps. We should not dismiss MEU because it does not match with peoples' actions anymore than Euclid was dismissed before triangulation. Rather we should turn our attention to the difficult problem of measurement of probability and utility. (Psychologists please note.) Perhaps it cannot be done. If so, an alternative theory will be needed. But surely it is premature to do it now. There are too many cases where MEU works for it to be superseded at the moment.

4. SMALL WORLDS

Savage's discussion of this topic is opaque and Shafer's attempt to clarify the matter is most welcome and his omelet example is marvelous. Here is an alternative way that I find useful for appreciating the very real difficulty exposed by Shafer. Savage's discussion is in terms of states s and consequences c; an act being a map from s to c. Another approach still uses states, which I prefer to denote by Greek letters, here using θ, and acts (or decisions) d. A consequence is then the ordered pair (d, θ). (Many other writers use this formulation or minor variants thereof.) A small world of d and θ can be enlarged by including another quantity ϕ in a more detailed state specification. So now the states, t in Shafer's notation, are pairs (θ, ϕ). The decisions are unaffected and the tortuous representation of consequences c (or (d, θ)) as acts mapping t to the new consequences (d, θ, ϕ) is avoided. In the omelet example θ is the state of the sixth egg, taking the values good or rotten (Table 1); ϕ is the state of the five eggs that have already been broken, taking the values fresh or stale (Table 9). The two tables have the same rows (decisions) but different columns (states) corresponding to the added refinement of the state of the five broken eggs.

In the (d, θ) description it is easy to see the relationships between the small and large world probabilities and utilities:

$$p(\theta) = \sum_{\phi} p(\theta, \phi) \tag{4.1}$$

and

$$u(d, \theta) = \sum_{\phi} u(d, \theta, \phi)P(\phi \mid \theta). \tag{4.2}$$

(Here, as in the example, ϕ assumes a finite number of values.) The difficulty with small and large worlds is that the small world assessment of $p(\theta)$ and $u(d, \theta)$ may not agree with the large world assessment of $p(\theta, \phi)$ and $u(d, \theta, \phi)$ according to these formulae. Omelets provide an example. (In the discussion I follow Shafer and suppose the washing of the saucer or the discarding of the egg do not enter into the utility, so that the consequences refer only to the state of the omelet; zero, five, or six eggs; Nero Wolfe or not).

In the small world it is tempting to say the consequences (d, θ) described by throw away/good and break into saucer/rotten have the same utility, since both result in a five-egg omelet. Now take the utilities $u(d, \theta, \phi)$ and probabilities $p(\theta, \phi)$ in the large world of Table 10 and calculate using (4.2). We easily have

$$u\text{ (throw away, good)} = 16(3/4) + 8(1/4) = 14,$$

and

$$u\text{ (break into saucer, rotten)} = 16(1/2) + 8(1/2) = 12,$$

so that they are not equal. There is thus a discrepancy between the small and large world views. It arises because θ and ϕ are not independent, a good egg having higher probability (3/4) when the others are fresh than when they are stale (1/4). Discrepancy could have arisen through the large world utilities but here only the probability causes trouble.

What is happening here is that consideration of a new feature (ϕ, the state of the five eggs) has changed your perception in the original small world. This is a common occurrence: "Goodness, I never thought of that." In its most extreme form we might just consider the decisions, assess their expected utilities, forgetting θ at all. We can enlarge by introducing θ, then further with ϕ, and so on until everything is included and we have Savage's truly large world. We presumably introduce θ because to do so will improve our decision making (whatever that means). Won't ϕ improve it further, and everything be better still?

A way of handling this genuine difficulty is to suppose that there are normative probabilities P and utilities U, and that the probabilities p and utilities u discussed above are measurements, subject to error, of them. A calculus of assessment errors (rather like least squares in triangulation) can be developed relating the lower and upper case values. Hopefully the introduction of ϕ will reduce the errors but at the cost of extra thinking. This is attractive because we have an MEU method of handling assessment errors in MEU; no new calculus is demanded.

5. ACTS

Shafer queries whether preferences among acts is really the basic idea. Many people have thought so. T. H. Huxley said, "The great end of life is not knowledge, but action." I agree with him. Action is all we have to go by. Why should we believe someone when they assert a probability of 0.8 or a utility of 12? But when they act, we can see them act, and ordinarily no doubts linger. Incidentally, this is one reason why I prefer the (d, θ) approach to that based on (s, c); decisions are primary, not derived as $f(s) = c$. It is a minor criticism of a stimulating paper that no mention is made of alternative axiomatizations, especially that of de Finetti whom Savage came to admire so much.

Comment

A. P. Dawid

I welcome Professor Shafer's interesting and thoughtful paper, not least for the stimulus it has given me to rediscover Savage's fascinating book and to ponder more deeply the place of axiomatic principles in statistics. I agree with much of Shafer's explicit criticism of Savage's work, but am not moved by his implied conclusion that the principle of maximizing expected utility needs modification.

THE NEED FOR AXIOMS

In his Preface to the Dover edition, Savage stated, "I would now supplement the line of argument centering around a system of postulates by other less formal approaches, each convincing in its own way, that converge to the general conclusion that personal (or subjective) probability is a good key, and the best yet known, to all our valid ideas about the applications of probability." This undogmatic, incremental approach to becoming a "Bayesian" describes well my own personal progress, and nails the axiomatic approach in place as one plank among many that form the Bayesian platform. Other arguments that have helped to sway me include: complete class theorems in decision theory; the quite distinct axiomatic approach via the likelihood principle (Berger and Wolpert, 1984); the unique success of de Finetti's concept of exchangeability in explaining the behavior of relative frequencies and the meaning of statistical models (Dawid, 1985a); the logical consequence of the Neyman-Pearson lemma that hypothesis tests in different experiments should use the identical indifference value for the likelihood ratio statistic (Pitman, 1965); the internal consistency of a Bayesian approach, in contrast to the many unresolved inconsistencies of every other approach; the conceptual directness and simplicity of the Bayesian approach in many otherwise problematic cases, both highly theoretical (as in asymptotic inference for stochastic processes; Heyde and Johnstone, 1979) and more applied (as in the calibration problem; Brown, 1982); and the general success of Bayesian methodology in the many practical situations to which it has been applied (Dawid and Smith, 1983).

Above all, I have adopted the Bayesian approach because I find that it yields the most fruitful insights into almost every statistical problem I meet. This is not to belittle the insights that other approaches may throw up, although these can usually be further illuminated by a Bayesian spotlight; nor would I claim total success in understanding, from any standpoint, such conundra as the role of experimental randomization, or the principles which should underly model criticism (Box, 1980). I even believe (and believe I have proved, Dawid, 1985b) that no approach to statistical inference, Bayesian or not, can ever be entirely satisfactory. I do, however, currently feel that the Bayesian approach is the best we have or are likely to have.

The trouble with relying only on axiomatic arguments is that they stand or fall according as one finds their postulates intuitively acceptable or not. I will often have strong feelings that a particular postulate or principle is, or is not, intuitively obvious, or acceptable, or inevitable; but I find that these feelings are not universally shared, and I generally cannot easily turn my gut feelings into arguments that will move dissenters. (They may be equally exasperated by my refusal to see reason.) That is why we should not attach too much importance to any axiomatic development such as Savage's, nor to Shafer's arguments

A. P. Dawid is Professor of Probability and Statistics and Head of the Department of Statistical Science, University College, London, Gower Street, London WC1E 6BT, England.