Discussion of "Scoring rules and the inevitability of probability, International Statistical Review 50, 1982, pp. 1-26.

Discussion of paper by D.V. Lindley

Glenn Shafer

Department of Mathematics, University of Kansas, Lawrence, Kansas 66045, USA

Congratulations to Dennis Lindley for this elegant and enlightening paper, which greatly clarifies the relationship between scoring rules and the Bayesian probability calculus. The message of the paper seems to be that the mere unidimensionality of a scoring rule suffices to encourage a person being scored to give numbers that correspond to additive probabilities; the special properties of 'proper scoring rules' are needed only if we want these numbers themselves, rather than transforms of them, to be additive probabilities.

Unfortunately, Lindley draws a further conclusion from his results: the conclusion that 'only probabilistic descriptions of uncertainty are reasonable'. This conclusion is unjustified. It is based on the blurring of the distinction between two very different tasks: the task of describing one's uncertainty, and the task of choosing numbers to minimize penalties incurred under a unidimensional scoring rule.

Consider the following artificial situation. A person is about to flip a coin. He knows that its true chance θ of coming up heads is between $\frac{1}{2}$ and $\frac{2}{3}$, but beyond this he has no further knowledge or evidence as to the value of θ or as to whether the coin will come up heads. In this situation it would be sensible, surely, for the person to use the numbers $\frac{1}{2}$ and $\frac{2}{3}$ to 'describe his uncertainty' about whether the coin will come up heads. Indeed, he might say that he has a degree of belief of $\frac{1}{2}$ that the coin will come up heads and a degree of belief of $(1-\frac{2}{3}) = \frac{1}{3}$ that it will come up tails. These numbers can even be given a betting interpretation; $\frac{1}{2}:\frac{1}{2}$ are the greatest odds at which the person knows a bet on heads to be at least fair, and $\frac{1}{3}:\frac{2}{3}$ are the greatest odds at which he knows a bet on tails to be at least fair. Now suppose the person is asked to give a number x for heads and a number y for tails, with the understanding that these numbers are to be scored by a quadratic scoring rule: he is to be penalized $(1-x)^2+y^2$ if the coin comes up heads and $x^2+(1-y)^2$ if it comes up tails. Under these circumstances, the person clearly should not give $x = \frac{1}{2}$ and $y = \frac{1}{3}$; instead he should choose some number p between $\frac{1}{2}$ and $\frac{2}{3}$ and give $x = p$ and $y = 1-p$. But it would not be accurate to call the number p he chooses a 'description of his uncertainty'; p involves not only his uncertainty but also an entirely arbitrary choice.

It seems unlikely that a person would ever be in the situation outlined in the preceding paragraph. But it is quite plausible that there are situations where one's uncertainty is better described by saying that the evidence is *like* knowing a chance to be at least p and at most q than by saying that the evidence is *like* knowing a chance to be exactly p. And this justifies an interest in the 'upper and lower probabilities' mentioned by Lindley in comment 11 of §5 of his paper. Other situations, where our evidence seems comparable to knowledge of chances governing the reliability and meaning of a message, similarly justify an interest in the 'belief functions' mentioned by Lindley in comment 12 of his paper. (For a discussion of the differences between upper and lower probabilities and belief functions, see my 'Constructive probability', in *Synthese* (1981) **48**, 1–60.)

After showing, in comment 11 of his paper, that the two degrees of freedom permitted by upper and lower probabilities cannot be used effectively in choosing numbers to minimize scores from a unidimensional scoring rules, Lindley writes as follows.

> This does not close the book on the idea of using two or more numbers to describe uncertainty, for it might be reasonable to use two or more score functions, measuring different qualities of the description in the manner of a multiattribute utility function.

The idea of scoring upper and lower probabilties on more than one dimension is an intriguing one; perhaps a person who gives the lower probability p and the upper probability q for an event E should be penalized both for his lack of certainty (measured perhaps by the minimum of q and $1-p$) and for the degree he doubts the truth (measured by $1-q$ if E happens, by p if $\bar{E}$ happens). But it is not clear what value such two-dimensional scores would have. And Lindley is wrong to suggest that the usefulness of upper and lower probabilities for describing uncertainty depends on whether they can be related in any interesting way to such scores.

A similar comment must be made on Lindley's discussion, in comment 12 of his paper, of belief functions. The values of belief functions are indeed intended as descriptions of uncertainty. But this does not mean that they are intended as input for any scoring scheme, unidimensional or multidimensional.

Lindley does not give any argument for his opinion that a set of numbers, in order to be a reasonable description of uncertainty, needs to be sensible as input for a scoring rule. But if he were called upon to defend this opinion, he would presumably give positivist arguments similar to those given by De Finetti in the volume he cites. According to these arguments, terms like 'probability' and 'description of uncertainty' are meaningless unless defined operationally, *i.e.* in terms of a person's economic choices. I believe these arguments are misguided. Once we recognize that real people do not have the consistent and thorough sets of preferences demanded by the Bayesian theory, simple-minded operational definitions lose their luster. Our real task is to construct probability judgments, not to define them in terms of nonexistent preferences. And in this constructive task we cannot afford a narrow positivism.

[*Received August* 1981]

Discussion of paper by D.V. Lindley

B.W. Silverman

School of Mathematics, University of Bath, Claverton Down, Bath, BA2 7AY, England

I should first of all like to congratulate Professor Lindley on a most interesting and stimulating paper. The basic message that probability is, in most operational senses, the only admissible description of uncertainty, seems very difficult to counter. We as statisticians are likely to be scored on our predictions whether we like it or not, and Lindley's elegant arguments deserve serious and careful consideration in this light.

It is perhaps a little disappointing, in some quarters at least, that the probabilities only turn out to be finitely additive. Of course this is hardly surprising in view of the fact that we are only ever asked to choose uncertainties for finite collections of events E_i. One is led to wonder whether insisting that infinite sequences of choices must also be admissible in an analogous way to (1) will lead to *countably* additive probabilities being the only admissible choices. The obvious way to calculate a total score corresponding to an infinite sequence of scores is to use a suitable discounted sum.

Suppose, therefore, that we were to impose the additional assumption of *countable discounted admissibility* as follows, for some fixed α, $0 < \alpha \leqslant 1$.

Given an infinite sequence of pairs (E_i, F_i) a person will *not* choose values x_i for E_i conditional on F_i $(i = 1, 2, \ldots)$ if there exist values $y_1, y_2, \ldots$ such that

$$\sum_i \alpha^i f(y_i, E_i) F_i \leqslant \sum_i \alpha^i f(x_i, E_i) F_i \qquad (1)$$

for *all* values of the indicator variables, and strict inequality holds for *some* values.