NOTES AND COMMENTS

On the Principle of Total Evidence

AYER (1957) raised the question of why, in the theory of logical probability (credibility), we should bother to make new observations. His question was not adequately answered in the interesting discussion that followed, by D. Bohm, R. B. Braithwaite, A. J. B. Cruikshank, P. K. Feyerabend, M. Fierz, W. B. Gallie, E. H. Hutten, W. C. Kneale, P. T. Landsberg, U. Öpik, M. Polanyi, L. Rosenfeld, M. Scriven, G. Süssmann, H. A. Thurston, and J.-P. Vigier. The question raised by Ayer is related by him to a principle called by Carnap (1947), ' the principle of total evidence ', which is the recommendation to use all the available evidence when estimating a probability. Ayer's problem is equally relevant to the theory of subjective probability, although, as he points out, it is hardly relevant to the theory of probability in the frequency sense.

In this note, Ayer's problem will be resolved in terms of the principle of rationality, the recommendation to maximise expected utility. (We use the words ' expected ' and ' expectation ' in the sense that is customary in nearly all books on mathematical probability or statistics.)

Our conclusion is that, in expectation, it pays to take into account further evidence, provided that the cost of collecting and using this evidence, although positive, can be ignored. In particular, we should use all the evidence *already* available, provided that the cost of doing so is negligible. With this proviso then, the principle of total evidence follows from the principle of rationality.

Suppose that we have r mutually exclusive and exhaustive hypotheses, $H_1, H_2, \ldots, H_r$, and a choice of s acts, or classes of acts, $A_1, A_2, \ldots, A_s$. It will be assumed that none of these classes of acts consists of a perpetual examination of the results of experiments, without ever deciding which of $A_1, A_2, \ldots, A_s$ to perform. Let the (expected) utility of act A_j if H_i is true be $U(A_j|H_i) = u_{ij}$. Suppose that, on some evidence, E, we have initial (prior) probabilities, $p_i = P(H_i|E)$. If just this evidence is taken into account, then the (expected) utility of act A_j is $\Sigma_i p_i u_{ij}$ and the principle of rationality recommends the choice $j = j_0$, the value of j that maximises this expression; and therefore the (expected) utility in the rational use of E is

$$\max_j(\Sigma_i p_i u_{ij}) = \Sigma_i p_i u_{ij_0} \quad .$$

We now consider making an observation whose possible outcomes are $E_1, E_2, \ldots, E_t$, where $P(E_k|H_i) = p_{ik}$ $(i = 1, 2, \ldots, r;\ k = 1, 2, \ldots, t)$. Let

$$q_{ik} = P(H_i|E.E_k) = p_i p_{ik} / \Sigma_i p_i p_{ik},$$

the final (posterior) probability of H_i if E_k occurs. (We denote logical conjunction by a full stop or period.) If in fact E_k occurs, then the expected utility of the use of E_k combined with E becomes

$$\max_j(\Sigma_i q_{ik}u_{ij}).$$

Now the initial probability of E_k is $\Sigma_i p_i p_{ik}$, so that the expected utility, in deciding both to make the new observation and to use it, is

$$\Sigma_k(\Sigma_i p_i p_{ik})\max_j(\Sigma_i q_{ik}u_{ij}) = \Sigma_k \max_j \Sigma_i p_i p_{ik} u_{ij}.$$

Accordingly we should like to prove that

$$\Sigma_k \max_j \Sigma_i p_i p_{ik} u_{ij} \geqslant \max_j \Sigma_i p_i u_{ij}, \tag{1}$$

with strict inequality unless the act recommended by the principle of rationality is the same irrespective of which of the events E_k occurs; in other words unless there is a value of j, mathematically independent of k, that maximises $U(A_j|E.E_k) = \Sigma_i q_{ik}u_{ij} = \Sigma_i p_i p_{ik} u_{ij}/\Sigma_i p_i p_{ik}$, or equivalently that maximises $\Sigma_i p_i p_{ik} u_{ij}$.

Since $\Sigma_k p_{ik} = 1$, the above proposition follows from the following Lemma by putting $f(j,k) = \Sigma_i p_i p_{ik} u_{ij}$.

LEMMA. *Let $f(j,k)$ be any real function of j and k. Then*

$$\Sigma_k \max_j f(j,k) \geqslant \max_j \Sigma_k f(j,k)$$

with strict inequality unless the matrix $\{f(j,k)\}$ has a 'dominating row'. (By a 'dominating row' of a matrix we mean a row in which each element is at least as large as any element in its own column.)

Proof of Lemma. Let a value of j that maximises $\Sigma_k f(j,k)$ be j_0. Clearly $\max_j(j,k) \geqslant f(j_0,k)$, since this would be true *however* j_0 were defined. The inequality is strict, when the definition of j_0 *is* used, unless $f(j,k)$ and $\Sigma_k f(j,k)$ are maximised by the same value of j. Therefore

$$\Sigma_k \max_j f(j,k) \geqslant \Sigma_k f(j_0,k) = \max_j \Sigma_k f(j,k).$$

This inequality is strict unless, for all k, $f(j,k)$ and $\Sigma_k f(j,k)$ are maximised at the same value of j. This establishes the Lemma and hence completes the resolution of Ayer's problem in terms of the principle of rationality.

At this point an opponent might say 'You have justified the decision to make new observations and to use them for the choice of the act A_j, but you have not justified the use of all observations that have already been made'. To this we can reply, 'The observations already made can be regarded as constituting a record. The process of consulting this record is itself a special kind of observation. We have justified the decision to make this observation and to use it, provided that the cost is negligible. In other words we have justified the use of all the observations that have been made, and this is the principle of total evidence.'

Our opponent might then say 'What you have shown is that, when faced with the two following possibilities, it is rational to select the second one:

(i) Not make an observation;

(ii) To make the observation and to use it for the choice of A_j;

but you have ignored a third possibility, namely

(iii) Make the observation and then not use it.'

My reply would be 'if we make an observation and then do not use it, this is equivalent to putting it back into the record. We have shown that it would then be irrational to decide to leave the observation in the record and not to use it, since there is a better course of action, namely to take it out (observe the record) and use it. You will now suggest other possibilities, such as the making of an observation, putting it on record, taking it out, putting it back, and so on, several times, and finally not using it. Our previous argument, with an obvious modification, shows that any such procedure is irrational, and it remains for you to suggest that your vacillating procedure should be continued for ever. But this would be a perpetual examination of the results of experiments, without a decision, and we have ruled this out by an explicit assumption.'

The simple mathematical theorem of the present note is not entirely new. Raiffa and Schlaifer (1961), p. 89, refer to the expected value of sample information, and seem implicitly to take for granted that it is positive. Lindley (1965), p. 66, explicitly states part of the theorem without proof. Perhaps the main value of the present note is that it makes explicit the connection between Carnap's principle of total evidence and the principle of rationality, a connection that was overlooked by seventeen distinguished philosophers of science.

Trinity College, Oxford I. J. Good

References

Ayer, A. J. (1957), 'The conception of probability as a logical relation', in *Observation and Interpretation*, ed. by S. Körner. London: Butterworths. Pages 12-30, including discussion

Carnap, R. (1947), 'On the application of inductive logic', *Philosophy and Phenomenological Research*, **8**, 133-148

Good, I. J. (1950), *Probability and the Weighing of Evidence*. London: Griffin

Good, I. J. (1960), 'Weight of evidence, corroboration, explanatory power, information and the utility of experiments', *J. Roy. Statistical Soc.*, B, **22**, 319-331

Lindley, D. V. (1965), *Introduction to Probability and Statistics*, Part 2. Cambridge: University Press

Raiffa, H. and Schlaifer (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University