Bayesian reasoning in science

Colin Howson and Peter Urbach

Bayesian scientific reasoning has a sound foundation in logic and provides a unified approach to the evaluation of deterministic and statistical theories, unlike its main rivals.

OURS is an uncertain world, though fortunately all things are not equally uncertain. But we are used to grading uncertainty. We are fairly certain, for example, that the textbook 'laws' of physics will remain valid for the foreseeable future, and very certain that they will remain in force through next week. We are much less certain about tomorrow's weather. We not only grade uncertainty — sometimes we measure it numerically, as in effect we do when we talk about the odds we think are merited by some predictive hypothesis or other. The more certain we are that an event will or will not take place, the longer (bigger) or shorter (smaller) the odds we are prepared to give.

Gamblers use the odds scale because odds tell you directly the proportions in which the stakes are divided after the outcome. But it is not a very good scale on which to measure uncertainty as such. Because odds are ratios the odds scale starts at 0 and is unbounded to the right (infinite odds). An equally balanced uncertainty, corresponding to 1 on the odds scale, therefore cannot be represented as a midpoint. To symmetrize, we transform odds into the scale of probabilities, using the formula probability (p) = odds/(1 + odds), and put the probability corresponding to infinite odds equal to 1. So probabilities lie between 0 and 1 inclusive, and even-money odds now become the midpoint of the probability scale, as desired.

Axioms

The formal theory of probability was born in the late seventeenth century in the work of Fermat, Pascal, Huyghens and James Bernoulli. It is summarized today in four basic laws, or axioms. The first says that the probability of any hypothesis h is a non-negative real number: $P(\mathrm{h}) \geqq 0$. The second says that the probability of any necessary truth t is 1: $P(\mathrm{t}) = 1$ (a necessary truth is one that is true whatever the world might be like; 'if it is raining then it is raining' is an example). The third, 'the additivity principle', says that if h and h′ are mutually exclusive then the sum of their probabilities equals the probability of their disjunction: in symbols, $P(\mathrm{h}) + P(\mathrm{h}') = P(\mathrm{h} \text{ or } \mathrm{h}')$. The fourth, says that the conditional probability $P(\mathrm{h} \mid \mathrm{h}')$ of h given h′, is equal to the unconditional probability $P(\mathrm{h}\&\mathrm{h}')$ of the conjunction h and h′, divided by the unconditional probability $P(\mathrm{h}')$ of h′ where that probability is positive: in symbols, $P(\mathrm{h} \mid \mathrm{h}') = P(\mathrm{h}\&\mathrm{h}') / P(\mathrm{h}')$, where $P(\mathrm{h}') > 0$.

Mathematical probability theory began life as theory of uncertainty. But in the late nineteenth century the probability axioms became recognized also as the laws of objectively random phenomena, or objective chance. Chance turns out to play an essential role in modern science, in the theory of statistical sampling, information theory, demography, genetics, thermodynamics and quantum theory.

Our concern here is with the older idea of probability as the foundation of a theory of uncertainty. We shall show how some modern developments endorse this idea, and how they enable us to see the rules of the probability calculus as a logic of inductive inference. Suppose h is some scientific hypothesis. Experimental data can never conclusively prove that h is true, even if it is true. So you are never absolutely certain of h's truth, only more or less. The inductive inference consists in assessing the degree of certainty warranted by the evidence. To the heirs of Bernoulli and Laplace, this means measuring the probability of h relative to data e. Scientists often estimate the probabilities of theories, but attempts to provide an objective basis for these estimates have uniformly failed.

But if the probability of a hypothesis merely reflects our own personal degree of belief in h, how can an objective logic of inductive inference be based on such probabilities? There is no paradox here. Your degrees of belief may be personal to you, but it does not follow that they are necessarily unprincipled or anarchic — in the first place, they must satisfy the axioms of probability. Of the many arguments that have been advanced to demonstrate this, the simplest is due to Frank Ramsey and Bruno de Finetti, who discovered it independently in the 1920s and 30s.

Their result is often called the Dutch book theorem. Consider a contract whereby one party agrees to exchange with the other a sum pS for the chance of receiving S if h is true and nothing if h is false. S is a non-zero sum of money or some other divisible good, called the stake. The payoffs to the first party thus are $S-pS$ if h is true and $-pS$ if h is false (the payoffs to the other party are of course the same with the signs reversed). The contract is tantamount to a bet in which the first party is betting on h at odds $pS{:}(S-pS)$, that is, odds $p{:}(1-p)$. It is easy to see that p is the quantity obtained by symmetrizing the odds through the transformation p = odds/(1 + odds), we introduced earlier. There we called p a probability, but so as not to prejudice matters we shall now call it the betting quotient associated with the odds.

Suppose p is such that you deem the odds $p{:}(1-p)$ fair, in the sense that to the best of your knowledge there is no advantage to taking either side of the bet. It is customary to identify this value of p with your degree of belief in h. Now consider some arbitrary finite set of hypotheses h_i, where p_i are your corresponding fair betting quotients on the h_i. A betting strategy with respect to the h_i is a set of decisions of the form 'bet on (against) h_i', for each i. Ramsey and de Finetti showed that if the p_i do not satisfy the probability axioms, then there are stakes S_i and a betting strategy for the h_i which must result in a certain loss for whoever follows that strategy (such a set of stakes is known as a Dutch book). Hence you cannot consistently maintain that the p_i are all fair if they do not satisfy the probability axioms.

Proof

The proof of the Ramsey-de Finetti result uses no more than high-school algebra. Consider axiom 2 of the probability calculus, that $P(\mathrm{t}) = 1$ if t is a necessary truth. Suppose that $p = P(\mathrm{t})$ is greater than 1. Because t is necessarily true the bettor on t will make a guaranteed loss of $Sp-S$. If p is less than 1 then the bettor against t will make a guaranteed loss of $S-Sp$. Either way one party or the other is guaranteed to lose, and so no value for p other than 1 can be fair. It is equally simple to see why $P(\mathrm{h})$ must be non-negative, where h is any hypothesis, and only slightly less straightforward to see how to justify the remaining two axioms of the probability calculus.

An immediate consequence of the probability axioms is Bayes's theorem, which has given its name to this approach. Thomas Bayes (1702–1761) was an English Nonconformist clergyman, a gifted mathematician, and a fellow of the Royal Society. His seminal work on probability is contained in one short memoir published posthumously in 1763.

Bayes's theorem says that, for any propositions h and e

$$P(\mathrm{h} \mid \mathrm{e}) = \frac{P(\mathrm{e} \mid \mathrm{h})P(\mathrm{h})}{P(\mathrm{e})} \qquad (1)$$

In the usual applications of the theorem, h is some hypothesis and e the evidence against which it is to be evaluated. $P(\mathrm{h} \mid \mathrm{e})$ is the posterior probability of h on e, $P(\mathrm{h})$ is the prior probability of h, and $P(\mathrm{e} \mid \mathrm{h})$ is the likelihood of h on e. Equation 1 can be rewritten as:

$$P(\mathrm{h} \mid \mathrm{e}) \propto P(\mathrm{e} \mid \mathrm{h})P(\mathrm{h})$$

That is, posterior probability is proportional to prior probability multiplied by likelihood. $P(e \mid h)$ can be interpreted as the degree of determinateness with which h explains e, and $P(h)$ is your estimate of the weight of evidence in favour of h before you know e. In practice it is by balancing empirical and prior factors in just this sort of way that hypotheses are evaluated. Neither factor by itself is decisive: the empirical success may, we might feel, be merely freakish, whereas the prior plausibility needs to be checked against empirical tests.

Elements

The probability calculus does more than explain how states of belief decompose into *a priori* and empirical elements. It also shows how the factor $P(e \mid h) / P(e)$ in Bayes's theorem, can be analysed to explain finer structure present in our informal reasoning. To see how, we need to invoke a further simple consequence of the probability axioms; namely, that

$$P(e) = P(e \mid h)P(h) + P(e \mid {\sim}h)\, P({\sim}h)$$

where ~h is the negation, or denial, of h. Substituting in equation 1, we get

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e \mid h)P(h) + P(e \mid {\sim}h)P({\sim}h)} \quad (2)$$

Equation 2 tells us that $P(h \mid e)$ depends also on both the factors $P(e \mid h)$ and $P(e \mid {\sim}h)$, or how probable e would be were h true or false, respectively. This dependence is more clearly brought out in the following equivalent formulation

$$P(h \mid e) = \frac{P(h)}{P(h) + \dfrac{P(e \mid {\sim}h)P({\sim}h)}{P(e \mid h)}} \quad (3)$$

Equation 3 says that because $P({\sim}h) = 1 - P(h)$, the posterior probability of h on e depends only on $P(h)$ and on the size of the so-called likelihood ratio $P(e \mid {\sim}h)$: $P(e \mid h)$, and approaches 1 as this ratio approaches 0. It is not difficult to show that if h′ is any hypothesis implying ~h, that is h′ is a potential alternative to h as an explanation of e, then the factor $P(e \mid h)$ is an increasing function of $P(e \mid h')P(h')$: in other words, the degree of confidence one will repose in h on the basis of e decreases with the extent to which e is explained by any plausible alternative to h. This is of course exactly the informal criterion used in practice for a strong positive evaluation of any hypothesis: no hypothesis is ever regarded as secure until there is no plausible alternative explanation of the data.

Scientific methodologies have to take a view about the kinds of conclusion scientists draw, and should provide appropriate mechanisms for those inferences. The bayesian method meets these conditions by characterizing a scientific conclusion about a hypothesis as a statement of its probability, and by providing Bayes's theorem as the mechanism for calculating that probability.

But bayesianism has been widely criticized because it is based on personal, hence subjective, probabilities. Scientific inference, critics say, should be perfectly objective. As one leading antibayesian philosopher put it: "The cognitive value of a theory has nothing to do with its psychological influence on people's minds . . . [but] depends only on what objective support it has in facts".

The objectivist ideal implicit in such criticisms is immensely attractive; it would be nice if disagreements in science could be resolved by taking the contending hypotheses and impartially measuring their 'objective cognitive values'. But is there any such thing?

Karl Popper and R. A. Fisher are among those who have tried to devise a purely objective methodology. Popper's starting point was the fact that general, deterministic hypotheses (simple example: 'All swans are white') can often be decisively falsified by evidence (for example, 'this is a black swan'). And his famous thesis is that only hypotheses that are falsifiable by possible or conceivable observations are scientific. A scientific theory may also make predictions which can then be experimentally checked. If the predictions are verified, Popper calls the theory corroborated. Now this is a perfectly objective statement about the theory, but does it amount to an evaluation, or carry information about the theory's cognitive value? Popper acknowledges that the hypothesis is not conclusively proved in the corroboration process; nor can it be said that its objective probability is augmented, because, as we shall explain, no one has managed to make sense of the notion of a theory's objective probability. How then can one interpret the corroboration notion without regarding it, Bayes-like, as reflecting our enhanced subjective belief in the corroborated hypothesis? Popper sometimes says that it is rational to prefer a corroborated hypothesis over one that is not, on the grounds that it is better tested; but this turns out, disappointingly, to be just another way of saying that it is corroborated. The dismal fact is that, as Popper conceded, a theory's degree of corroboration is simply a record of its performance in the various tests, with no implication for its 'cognitive value'.

A statistical hypothesis attributes statistical probabilities, or chances, to events. These are not the subjective degree-of-belief probabilities discussed earlier, but objective properties of repeatable experiments. A simple example is the hypothesis that a particular coin is fair, that is, has equal statistical probabilities of a half of landing heads and tails. And what this is standardly taken to mean is that if you tossed the coin repeatedly, the relative frequency of heads in the resulting sequence of outcomes would tend to 0.5, as the number of throws increased to infinity.

Modern science deals extensively with statistical theories, so how they are tested and evaluated is an important question. But Popper's approach is inapplicable to statistical hypotheses, which are not falsifiable.

The question of how to test and evaluate statistical hypotheses in an objective, nonbayesian way was taken up by Fisher in the 1920s, and his approach has developed into an influential body of doctrine, known as the classical theory of statistical inference.

Classical statistical inference has two principal parts, the first relating to the testing of hypotheses (using signficance tests) and the second to estimating the values of unknown parameters. The essential principles of classical inference can be appreciated through the simplest examples. Consider again the hypothesis that this coin is fair. To test it, you might toss the coin a predetermined number of times, say 20, and count the number of heads obtained. There are 21 possibilities, ranging from no heads and 20 tails to 20 heads and no tails. For a significance test, the statistical probability of each, relative to the test, or null hypothesis, must be calculated. The classical method is then to choose a subset of the possible results — usually in one or both of the distribution's tails. Several considerations dictate this subset (or critical region); an important one is that the probability that any actual outcome of the experiment falls within it when the null hypothesis is true is fairly small; 0.05 has established itself as an acceptably small value. Finally, if the outcome obtained in the experiment is in the critical region, it is said to be 'significant at the 5 per cent level'.

Whether a result is significant at a particular level seems a perfectly objective fact. And classical statisticians believe that from it one can draw a similarly objective judgment about the null hypothesis. Hence, a result significant at the 5 per cent level is said to entitle one to 'reject the null hypothesis at the 5 per cent level'.

The meaning of this expression needs analysis, however, for although the idea of rejecting a hypothesis is fairly intuitive, the idea of doing so at some percentage level is not. Fisher sometimes wrote of the hypothesis being disproved in a significance test, though he of course knew that there can be no logical disproof of statistical hypotheses. Statisticians now agree that all one can infer for certain about a hypothesis confronted with a significant result is that either the hypothesis is true, in which case the result is relatively improbable, or it is false. Indeed, Fisher held that the "force of a test of significance" resided precisely in this dichotomy. But the dichotomy has no force, as it says nothing about the null hypothesis beyond the tautology that it is either true or false.

Advice

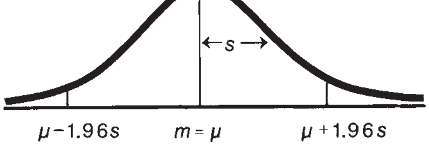
Another response to a significant result is due to Jerzy Neyman and Egon Pearson, who invented the currently standard form of a significance test, which is slightly more complicated than that given above, in that rivals to the null hypothesis are brought into the picture. Neyman and Pearson suggested that although we are not entitled to believe that the null hypothesis is false, we should act in our practical life as if we did believe that. This oft-repeated advice is always justified by saying that if you performed significance tests repeatedly and if each time the result was significant at the 5 per cent level you acted as if believing the null hypothesis was false, you would "in the long run" be wrong on only "around 5 per cent" of the occasions.

This argument, although superficially plausible, is mistaken in every respect. First, it is surely absurd to act as if you believed a hypothesis false, when you are uncertain that it really is. It would mean, for example, that you would willingly accept a wager in which you received £10 if the hypothesis was finally established to be false, and you had to forego all your wordly goods if it was finally established to be true. No reasonable person would accept such a bet, without being absolutely certain that the hypothesis in question really was false. Second, the justification is fallacious. It starts from the fact that the probability of rejecting a true null hypothesis equals the significance level; from this, it infers that if a test were repeatedly performed using a 0.05 significance level, the approximate frequency of rejecting a true null hypothesis would be 5 per cent, in the long run. But the inference is simply a nonsequitur: you cannot derive from the probability of an event even the approximate frequency with which that event will appear in any actual run of trials, however long. Finally, the justification is irrelevant, because it refers to the supposed frequency of drawing a wrong conclusion in a sequence of possibly imaginary tests, but says nothing about the particular case at hand.

Scientists often need to know the value of physical parameter that cannot be measured directly. The task of gauging the mean height of a very large population is a simple example. In such cases, an indirect measurement must be made and this is standardly done by examining a suitably selected sample.

Techniques

Classical statisticians have devised techniques that supposedly permit objective estimations of parameters, the principal one being that of the confidence interval, which we can explain through the example already mentioned. Let the unknown, mean population height be μ. Assume that we know the standard deviation, σ, of heights in the population. Now suppose a random sample of size n is drawn from the population. The mean height of people in such a sample is a random variable, m. Clearly m can take many possible values, some more probable than others. The distribution representing this situation is 'normal' and its standard deviation is given by $s = \sigma/\sqrt{n}$:



This distribution is a plot of possible sample means against probability densities, not probabilities. The important fact for the present discussion is that the probability that m lies between two points is proportional to the area enclosed by them and the curve. Because the distribution is normal, it follows that with probability 0.05, $-1.96s \leqq \mu - m \leqq 1.96s$. Rearranging these inequalities gives the result that with probability 0.95, $m - 1.96s \leqq \mu \leqq m + 1.96s$. Suppose m' is the value of m that is actually observed in the experimental sample. Then, because we know σ and n, the terms $m' + 1.96s$ and $m' + 1.96s$ can be determined; the interval between them is called a 95 per cent confidence interval for μ, and classical statisticians regard such an interval as a reasonable estimate of μ.

The statement that such-and-such is a 95 per cent confidence interval for μ seems objective. But what does it say? It might be imagined that a 95 per cent confidence interval corresponds to a 0.95 probability that the unknown parameter lies in the confidence range. But in the classical approach, μ is not a random variable, and so has no probability. Nevertheless, statisticians regularly say that one can be '95 per cent confident' that the parameter lies in the confidence interval. They never say why.

In fact, there is a decisive reason why not. The confidence interval is derived from the probability distribution of sample means depicted above. This distribution gives the probabilities of all the sample means that you might have got in the experiment. It is usual to assume that all those possible samples have the same size as the actual sample. This assumption is crucial, because the shape of the distribution, and hence the width of the confidence interval, is affected by that size. Now the set of possible samples is determined by the experimenter's intention. If he had deliberately set out to sample exactly n people, the usual assumption would be justified. But suppose each time he selected a person from the population, he also tossed a fair coin, and that he planned to stop sampling as soon as the coin had produced 5 heads; or suppose the plan was simply to examine as many people as possible before lunch, or before getting bored. With any of these plans, the experimenter might still have arrived at a sample of n, but the set of possible samples would have been different, and hence, so would the confidence interval. So the degree of confidence we are invited to place in an estimate inevitably depends on the private plans of the experimenter, which is surely immensely counterintuitive.

This is the so-called stopping-rule problem. It also affects significance tests. In our earlier example, it was assumed, as it normally would be, that because the coin was tossed 20 times, all the of possible outcomes would exhibit 20 heads and/or tails. But these are the possible outcomes only if the experimenter had a premeditated plan to throw the coin 20 times. Had the plan been to stop the experiment when, say six heads appeared, he could have got just the result he did, but with a different list of unrealized, possible outcomes. Because significance is calculated by reference to these possibilities, a result could be significant if the experimenter had had one plan (or stopping rule) in mind, but not significant if it was another.

This dependence of significance tests and confidence interval estimates on the subjective, possibly unconscious intentions of the experimenter is an astonishing thing to discover at the heart of supposedly objective methodologies. It is also a most inappropriate thing to find in any methodology, for the plausibility, or cognitive value, of a hypothesis, and our rational confidence in an estimate should not depend on the contents of the experimenter's mind.

Illusion

Popper's corroboration idea, and the theories of significance tests and confidence intervals were developed as supposedly objective methodologies in conscious reaction to subjective bayesianism. These methods all issue in apparently objective statements, couched in a deceptive terminology which gives the impression that an important, objective, theoretical evaluation is being achieved. But this is illusory. Corroborating a hypothesis does not strengthen it, a significant result has no significance for the truth of the null hypothesis, and a 95 per cent confidence interval has no right to impart confidence, let alone 95 per cent's worth, to an estimate.

Unlike these pseudo-objective methodologies, the bayesian approach has a solid foundation. It provides a unified approach to deterministic and statistical theories, and to questions of testing and estimation, unlike the many *ad hoc* recipes of the classical approach. It is also intuitively right. We can illustrate this by sketching the bayesian way of estimating a population mean. It starts with a distribution of subjective prior probabilities over the range of possible values of the parameter. Then, using Bayes's theorem and the sample evidence, a corresponding posterior probability is calculated. The prior probability curve typically would be very spread out, indicating considerable initial uncertainty about the parameter value, whereas the posterior probability would be concentrated in a narrow region. Then if 95 per cent of the area under the curve was enclosed between two points a and b, the bayesian estimate of the parameter would be of the form 'μ lies between a and b with probability 0.95'.

This bayesian conclusion has a clear meaning and is just the kind of conclusion people do come to. It is derived from the mean of the experimental sample alone, not the means of possible samples. Hence, it is unaffected by the experimenter's subjectively intended stopping rule, which is as it should be. Finally, the posterior distribution is very insensitive to variations in the prior distribution, and this insensitivity increases rapidly with the sample size. Hence two people starting from very different prior beliefs must converge in their posterior beliefs as evidence accumulates. Again this seems realistic and shows that the objection to bayesianism that it makes scientific inference purely subjective is misguided.

A traditional complaint against the bayesian theory is that it introduces numerical calculations into areas where a rough and ready reliance on intuitive methodological principles seems to have worked well enough. Indeed, bayesian theory subsumes much basic methodological lore. Why go to all the trouble of writing down difficult formulas instead of continuing to do what we did well enough without them?

One answer is that 'writing down these formulas' exhibits intuitive procedures as consequences of fundamental principles of logic. Doing something correctly is one thing; to know you are doing it correctly, and why, is equally important. But there is a practical reason also. The rise of artificial intelligence and in particular the development of rule-based expert systems have made the question of what is the best method for modelling uncertain reasoning a matter of great practical urgency.

Bayesianism is only one of the mathematical theories of uncertainty currently under evaluation. The others were largely inspired by dissatisfaction with the bayesian approach. The principal alternatives are the Dempster–Shafer theory[1], incorporating a nonprobabilistic measure of belief due to Shafer, and Dempster's rule for combining evidence; 'possibility theory', put forward in the 1970s and based on fuzzy set-theory[2,3]; and the approach, also developed in the 1970s, based on so-called certainty factors and embodied in the MYCIN and EMYCIN calculi[4].

Uncertainty

All these theories embody numerical measures of uncertainty, and rules for revising uncertainties under the impact of new evidence. All are distinct from each other, though there turn out to be systematic relationships between probabilities, the belief and plausibility measures of the Dempster–Shafer theory, and possibility measures[3]. But only in the bayesian theory must the assignments of numerical uncertainty to hypotheses satisfy the probability axioms. Although in the Dempster-Shafer theory the fundamental quantities are called basic probability numbers, defined on all the subsets (roughly, hypotheses) of a frame of discernment (hypothesis space), these are not formally probabilities, nor is the belief function (Bel) constructed from them. In fact, all the other theories we have mentioned contrast with the bayesian by being nonadditive: that is, the quantities functioning as degrees of belief in each do not add over incompatible alternatives.

There is a vigorous and occasionally polemical debate under way on the relative merits of these different approaches to modelling uncertainty. With its profusion of manifestos, it recalls an earlier controversy about the role of logic in knowledge-based computer systems. One prominent defender of the bayesian view, Peter Cheeseman, has emphasized[5] the analogies with that earlier debate; logic, like probability, had historical precedence and was well understood, and the newer representation languages, like the newer theories of uncertainty, based their claims on alleged problems with a logic-based approach.

Controversy

Some of the controversy inevitably turns on the ease with which the various approaches can be implemented in computer systems; bayesian models are not always the most practical in these situations. They tend to have exponential informational complexity, as is also the case with Dempster–Shafer theory, depending on how many hypotheses are assigned non-zero probability. By contrast, MYCIN and EMYCIN have linear complexity. But the fundamental disagreements are philosophical, not practical, and one of the most controversial conceptual problems alleged to arise within the bayesian theory, is that of evaluating the probabilities, especially the prior probabilities, to be plugged into Bayes's theorem. In most applications there is relevant background information: in a disease diagnosis problem, for example, this would consist of the available clinical data, together with some theory. All this information should be made use of in determining probabilities. But how should this be done?

Bayes's theorem is the standard technique by which probabilities are adjusted to data (though there are generalizations of it which collectively go under the heading of probability kinematics), and the priors in any calculation can also be posterior probabilities calculated from the historical data. Ditto for the priors used in the calculation of those posteriors. But at some point in this backward progress, prior probabilities will have to be used which merely reflect opinion. This allegedly opens the doors to an unwelcome subjectivism.

To deflect the criticism, there have been several attempts to provide explicit, 'objective' rules for calculating priors. The best known is as follows: regard the hypotheses as defining subsets of some appropriate universe of 'elementary' possibilities, in the same sort of way that the hypothesis 'this die will land an even number upwards' defines the subset $\{2,4,6\}$ of the set $\{1,2,3,4,5,6\}$ of possible outcomes of throwing the die. Now make the prior probabilities, or probability densities if they form a continuum, of all these possibilities equal. If the universe of possibilities is finite and has m members, and if h is true in n of these, then the desired 'objective', or, as it is sometimes called, 'informationless' prior probability of h is equal to n/m.

A recent attempt to justify equal, or uniform, prior probability assignments appeals to the maximum entropy principle (MEP), according to which 'informationless' prior distributions are those which maximize uncertainty, in the sense of Shannon entropy. For in the absence of constraints other than those imposed by the probability axioms themselves, MEP prescribes uniform prior distributions.

A technical difficulty with uniform prior distributions is that they cannot properly be defined over infinite intervals, like the whole real line. But the major objection to this strategy is the arbitrariness implicit in the choice of the space of equiprobable alternatives. There are in general many different ways of embedding a set of hypotheses and data in some space or other of possibilities, and it turns out that depending on which you choose, you may get different prior probabilities for your hypotheses. And the choice is bound to be arbitrary because it is supposed to be made in advance of any empirical information (for a more detailed discussion of these matters see ref. 6).

In short, there seems to be no way of 'objectively' defining prior probabilities. But our argument has all along been that this is really no weakness: it allows expert opinion due weight, and is a candid admission of the personal element which is there in all scientific work. The inventors of 'objective' methodologies, as the statistician and philosopher I. J. Good is fond of remarking, merely sweep the personal element under the carpet. Also, there are various convergence-of-opinion theorems, which show that posterior distributions based on a lot of data are usually insensitive to priors.

The subjective bayesian theory is the only one of the various calculi of uncertainty so far to have been given an explicit and powerful theoretical justification; and the argument based on the Dutch book is only one of several. The other theories have all been proposed more or less *ad hoc*, in response to supposed difficulties in the bayesian scheme. These other approaches are not valueless; some can even be regarded as interesting generalizations of the bayesian theory. But if a clear and flexible calculus of undcertain reasoning is required, with a secure foundation in logic, then the subjective bayesian theory has no real rival. □

Colin Howson and Peter Urbach are in the Department of Philosophy, Logic and Scientific Method, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

1. Shafer, G. *A Mathematical Theory of Evidence* (Princeton Universiity Press, Princeton, 1976).
2. Zadeh, L. *Fuzzy Sets Systems* **1**, 3–28 (1978).
3. Dubois, D. & Prade, H. *Acta psychol.* **68**, 53–78 (1989).
4. Buchanon, B. G. & Shortliffe, E. H. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristics Programming Project* (Addison-Wesley, Reading, 1984).
5. Cheeseman, P. *Proc. 9th Int. Joint conf. AI, Los Angeles* 1002–1009 (1985).
6. Howson, C. & Urbach, P. M. *Scientific Reasoning: The Bayesian Approach* (Open Court, La Salle, 1989).