JONATHAN BARON

SECOND-ORDER PROBABILITIES AND BELIEF FUNCTIONS

ABSTRACT. A second-order probability $Q(P)$ may be understood as the probability that the true probability of something has the value P. "True" may be interpreted as the value that would be assigned if certain information were available, including information from reflection, calculation, other people, or ordinary evidence. A rule for combining evidence from two independent sources may be derived, if each source i provides a function $Q_i(P)$. Belief functions of the sort proposed by Shafer (1976) also provide a formula for combining independent evidence, Dempster's rule, and a way of representing ignorance of the sort that makes us unsure about the value of P. Dempster's rule is shown to be at best a special case of the rule derived in connection with second-order probabilities. Belief functions thus represent a restriction of a full Bayesian analysis.

1. INTRODUCTION

When we are asked the probability of a coin coming up heads, we do not hesitate to answer, "0.5", or "about 0.5". In contrast, when we are asked the probability of a Republican being elected president in 1996, we hesitate to assign a probability. Similarly, physicians familiar with the relevant statistics will willingly provide a statement about the probability of a white male who smokes two packs a day eventually developing lung cancer, but, faced with a particular white male smoker, many will hedge: "I can tell you only the probability in general; I cannot say what the probability is for you". People sometimes are uncertain about what their probability assignment ought to be.

Manipulation of such uncertainty in experiments (Ellsberg, 1961; Einhorn and Hogarth, 1985) indicates that people really behave differently when their best guess about a probability is a "firm" one than when it is not. For example, under conditions of "ambiguity", people often behave as if they had assigned high probabilities to bad outcomes, thus avoiding actions that might lead to such outcome even if these are the same actions that might lead to the best outcomes. When different outcomes are contingent on the same events, this tendency can lead to violations of the principle of independence, i.e., when there is a state of

Theory and Decision **23** (1987) 25–36.

the world that yields the same outcome regardless of one's choice, the outcome should not affect the choice made.[1] Ellsberg (1961) and others have defended such behavior despite such violations, on the ground that people are intuitively inclined to find it reasonable. An understanding of the nature of "ambiguity", from both normative and descriptive views, could clarify the behavior of subjects in these experiments and help to assess its rationality.

Uncertainty about probabilities creates difficulties when we seek to carry out a formal analysis of some judgment or decision and when the analysis requires subjective probabilities as inputs. For example, in the evaluation of risks of nuclear power plants, there is as yet no generally acceptable way to take into account our feeling that the true probability of accident is, in some important sense, unknown. Similarly, in the design of expert diagnostic systems, many workers have assumed that certain probabilities are unknowable.

This note concerns two proposed solutions to the problem of carrying out a formal analysis when judges are unwilling to state exact probabilities or when they feel that their probability assignments are themselves unreliable: a version of the idea of second-order probabilities (similar to one proposed by Raiffa, 1968) and belief functions (Shafer, 1976; Shafer and Tversky, 1985). I shall suggest that the method of belief functions is at best a special case of the method of second-order probabilities, and I shall comment on the translation between the two methods when it is possible. To make this comparison, I shall concentrate on the way in which two independent sources of evidence may be combined under each system.

2. SECOND ORDER PROBABILITIES

Let us suppose that there is a true probability P assigned to some proposition. We might also imagine having a subjective second-order probability distribution $Q(.)$ over the possible values of P. Here, we shall make the simplifying assumption that P can take on a finite number of different values, for example, 0.01, 0.02, ..., 1.00. We treat P as an unknown parameter, about which we have subjective beliefs just as if it were any other unknown parameter such as the population of New York City.[2] If we are very uncertain about the true value of P, $Q(.)$ will broadly spread out. If, on the other hand, we are certain of P, $Q(.)$ will be concentrated at a point: in the case of a fair coin, $Q(P)$ might be zero for all values of P except 0.50, where $Q(P)$ would be 1.

There are several things that we might take "the true probability (P)" to mean. For example, we might mean the probability that we would assign if our assignments were coherent (Lindley *et al.*, 1979). The arguments to be made concerning the relation between second-order probabilities and belief functions do not depend on the meaning that we choose.

Despite the fact that we do not need to settle on an interpretation of P, there is one interpretation (similar to that of Brown, 1986) that may be particularly useful in thinking about the question. In particular, we may think of P as *the probability that we would assign if we had access to a certain body of information.* The unknown information might include relevant population statistics or easily collected evidence such as test results. It might include the results of examining our own mind more carefully, either through rehearsal and contemplation of evidence or through the making of calculations (e.g., those suggested by Lindley *et al.*). It might include the assessments of other people, or additional data.

One advantage of this formulation is that it is highly general. It incorporates, for example, the concept of ambiguity: the missing evidence may consist of the proportion of white balls in an urn we are about to draw from, or which of two expert probability assessments is to be believed.

A second advantage is that it is clearly defined, at least compared to similar formulations (e.g., Gärdenfors and Sahlin, 1983). The clearer our idea of just what information is missing and how it might affect our beliefs, the more precisely we can specify $Q(.)$. In the case of missing medical-test data, we can specify $Q(.)$ with substantial precision.

A third advantage of this formulation is that it might be P, rather than our current best guess about P, that we are interested in, and we might well need to take into account the fact that certain information is lacking. Brown (1986) gives the example of meeting a legal requirement that the probability of a nuclear accident be below some specified number. Although our current best guess at P may be below the value, we may be able to imagine information that would push our P above the cutoff. In this case, we cannot guarantee that the legal requirement can be met.

If the information is obtainable, we may want to obtain it. This way of looking at second-order probabilities calls attention to their similarity to the idea of assessing the value of information itself by examining its potential impact on our probabilities and consequent decisions (Baron, 1985, ch. 4).

This formulation makes the concept of second-order probability dependent on a specification of which evidence is lacking. In principle, we might subscript $Q(.)$ for different bodies of evidence, although this will not be necessary here. This potential instability of $Q(.)$ may be considered as a disadvantage of the present formulation, but it may also be an advantage. Our degree of dissatisfaction with precise assessments of P seems to change as we focus our attention on different bodies of evidence. For example, we feel unease about assignment of probabilities to past events (even if unknown to us) because the information is potentially available. Likewise, we may acquire a feeling of ambiguity about the drawing of a card from a deck if we think that all but the top three cards have been written down on a piece of paper, which is now face down in front of us on the table. Moreover, in practical decision-analysis, the missing information that is relevant may differ from case to case.

Combining independent sources of evidence

Let us consider how we might use the idea of second-order probabilities to combine evidence from two sources concerning some proposition. We want to do this in order to compare the result to that from the use of belief functions. We want to combine two distributions, $Q_1(.)$ and $Q_2(.)$, with a prior distribution over the values of P, $p(.)$, in order to arrive at a final distribution $Q(.)$. The prior is assumed to be identical for all parties. For example, the proposition may be that a certain company will go bankrupt in the next year, and the evidence may consist of summary statements of two independent, equally credible, auditors. One auditor might say, "I'd give it a probability of 0.90, but I haven't seen the last three quarterly reports. If I did, I could imagine changing my probabilities to a substantial degree". The other might say, "I'd give it 0.55, but I've seen only the last three reports. If I saw the others, I could imagine changing a little". For each piece of evidence E_i, taken by itself, we have a subjective probability distribution $Q_i(P)$. If we take each auditor's judgments to be well calibrated (so that there is no systematic bias) we would expect that our own mean of $Q_1(.)$ is about 0.80 and our mean of $Q_2(.)$ is about 0.55. The variance (or some other appropriate measure) of each distribution will depend on how much we think the announced probability might change if the auditor had the missing reports in hand. When we combine the two pieces of evidence, our own personal mean will be closer to the mean of the second, because he had more information (other things being equal).

Let us assume that each distribution $Q_i(.)$ is the Bayesian posterior that results from combining E_i with the prior distribution $p(.)$. In particular, by Bayes' theorem,

$$Q_i(P) = p(P \mid E_i) = p(E_i \mid P) \cdot p(P)/p(E_i).$$

Similarly,

$$Q(P) = p(P \mid E_1 \& E_2) = \frac{p(E_1 \& E_2 \mid P) \cdot p(P)}{p(E_1 \& E_2)}.$$

Assuming conditional independence of E_1 and E_2 given each value of P, this implies that

$$Q(P) = \frac{p(E_1 \mid P) \cdot p(E_2 \mid P) \cdot p(P)}{\sum_P p(E_1 \mid P) \cdot p(E_2 \mid P) \cdot p(P)}.$$

The summation is over all the values of P under consideration. Since $p(E_i \mid P) = p(\mathrm{P} \mid E_i)p(E_i)/p(P)$,

$$Q(P) = \frac{\dfrac{p(P \mid E_1) \cdot p(E_1)}{p(P)} \cdot \dfrac{p(P \mid E_2) \cdot p(E_2)}{p(P)} \cdot p(P)}{\sum_P \dfrac{p(P \mid E_1) \cdot p(E_1)}{p(\mathrm{P})} \cdot \dfrac{p(P \mid E_2) \cdot p(E_2)}{p(P)} \cdot p(P)}.$$

Canceling $p(E_1)$ and $p(E_2)$, and replacing $p(P \mid E_i)$ with $Q_i(P)$, we get

$$Q(P) = \frac{Q_1(P) \cdot Q_2(P)/p(P)}{\sum_P Q_1(P) \cdot Q_2(P)/p(P)}.$$

If the prior $p(.)$ is constant over all possible values of P (as if would be if we have no evidence at all and if we invoke the principle of insufficient reason), $p(P)$ may be canceled as well, and we get

$$(1) \qquad Q(P) = \frac{Q_1(P) \cdot Q_2(P)}{\sum_P Q_1(P) \cdot Q_2(P)}.$$

We shall assume constant $p(P)$ when we compare the present model to belief functions. This amounts to the assumption that all parties start out with complete ignorance, so that each function $Q(.)$ represents all the evidence available to each party.

Equation 1 shows how $Q(P)$ may be calculated for each value of P by combining two sources of information (and assuming constant $p(P)$). It is apparent that the detailed shapes of $Q_1(.)$ and $Q_2(.)$, as well as their means and variances, will have an effect on the final distribution $Q(.)$.

It is important to note that the theory of second-order probabilities is fully consistent with Bayesian probability theory (at least if we interpret P as "the probability that would be assigned if a certain body of evidence were available"). Thus, this theory comes with both a justification in terms of necessary axioms (Savage, 1954) and some useful measurement procedures (Krantz *et al.*, 1971, chs. 5, 8; Raiffa, 1968).

3. BELIEF FUNCTIONS

Shafer's (1976) theory of belief functions also allows us to combine evidence in situations like the one we have been considering. This theory differs from probability theory in that it allows us to apportion our total belief among overlapping sets of propositions as well as mutually exclusive sets. For example, we may assign some belief to the set including all propositions under consideration and the rest of our belief to one proposition, a subset or member of that set. The assignment of belief to the former set is uncommitted. The amount of uncommitted belief might be taken to correspond to the amount of second-order uncertainty.

More formally, it is assumed that the total belief adds to one (as in probability theory). In the simplest situation in which two pieces of evidence are combined, no belief is committed to the falsity of the proposition, and both pieces of evidence support exactly the proposition in question (as opposed to some proposition that implies it, is implied by it, or is consistent with it) to some degree.[3] For each piece of evidence i, A_i is committed to the proposition in question and $1-A_i$ is uncommitted. Thus, each piece of evidence is fully characterized by A_i. We may combine the two pieces of evidence to produce a resulting belief, which may be characterized by A. Thus, we want A as a function of A_1 and A_2. This is given by Dempster's rule of combination, the crux of the theory. Applied to the present case, this rule specifies that the resulting belief committed to the proposition is

$$A = A_1 \cdot A_2 + A_1 \cdot (1 - A_2) + A_2 \cdot (1 - A_1)$$

or

$$(1 - A) = (1 - A_1) \cdot (1 - A_2). \quad (2)$$

Shafer gives no particular justification of this rule, although he notes that it corresponds to the following situation: Each piece of evidence E_i is like a coded message. There are several codes that you might use to decode the message. You do not know which code to use, but you do know that all the codes are equally likely to be correct. You try all the codes and discover that a proportion A_i of them decode the message to "The proposition is true" (e.g., "The company will go bankrupt"). The remaining codes, that is, proportion $1 - A_i$, decode to some utterly uninformative message, such as, "The proposition is either true or false". There is a different set of codes for each message. The value of A in equation 2 is thus the probability that either message, properly decoded, indicates that the proposition is true.

4. EQUIVALENCE OF SECOND-ORDER PROBABILITIES AND BELIEF FUNCTIONS

Suppose we have a rule for translating functions $Q_i(.)$ into values of A_i. If we combine two pieces of evidence according to equation 1 so as to get $Q(.)$, under what conditions will it be true that equation 2 holds as well, given the translation rule? Clearly, the rule will characterize any particular $Q(.)$ by a single parameter, B. (Let us use the generic form $Q(.)$ to stand for $Q(.)$ or $Q_i(.)$.) We seek, then, a function $A = g(B)$ such that equation 1 implies equation 2. It will be convenient to define $f(B) = 1 - g(B)$. Thus, from equation 2, $f(B) = f_1(B) \cdot f_2(B)$. (Note that $f(.)$ cannot be unique. If $f(.)$ satisfies the condition, so does $[f(.)]^r$ for any nonzero r.)

If we can find such a function $f(.)$ that applies to any distribution $Q(.)$, then the system of belief functions and the system of second-order probabilities are equivalent, at least for the simple sort of case under consideration. Alternatively, if this requirement places some limitation on the function $Q(.)$, then belief functions correspond to a special case of the theory of second-order probabilities (provided that there are functions that satisfy the limitation). In other words, B may be defined only for a subset of the possible distributions $Q(.)$. Finally, it may be the case that *no* reasonable distributions $Q(.)$ satisfies the condition. In this case, belief functions cannot be represented in terms of an extremely flexible model of second-order probability.

To begin, let us consider two second-order probability functions, $C(.)$ and $U(.)$ (for certain and uncertain, respectively), which must be included in the set of functions $Q(.)$ to which the translation rule applies.

$C(.)$ is the function $Q(.)$ when the subject is *certain* that the proposition is true. $C(.)$ must be concentrated at $P=1$, i.e., $C(P)=0$ for $P<1$, $C(P)=1$ for $P=1$. When $A=1$, it must also be the case that $P=1$, because any second-order uncertainty must make the mean of $Q(.)$ less than 1. Similarly, when $A<1$, it must be true that $Q(1)<1$ (for otherwise, P would be 1).

$U(.)$ is the function $Q(.)$ when the subject is completely *uncertain* and has absolutely no evidence, i.e., when $A=0$. It ought to be true that $U(.)=p(.)$. Absence of evidence should leave the prior, $p(.)$, unchanged. It follows that if $p(1)>0$ then $U(1)>0$. In the cases of interest, we have assumed that $p(P)$ is constant for all values of P.

For belief functions of the sort we are considering, in which there is no evidence against the proposition in question, it is always possible to achieve certainty. All that is required is to combine our current belief with another belief in which $A=1$. Hence, it must be the case that $Q(1)>0$ for any $Q(.)$. (If $Q_i(1)=0$, it would be impossible to make $Q(1)=1$ by combining $Q_i(.)$ with $C(.)$.) However, in second-order probabilities it might be reasonable that $Q(1)=0$. For example, in diagnosing a disease, the only additional evidence available might consist of fallible tests, none of which could possibly make the probability of the disease more than 0.90. $Q(1)$ would thus be 0, contradicting the conclusion just drawn that $Q(1)>0$. Thus, it is already apparent that the wish to translate between second-order probabilities and belief functions places constraints on the former, constraints that would not otherwise be present in many realistic cases.

It might be argued that such an upper bound on P implies that we have some evidence against the disease, contrary to our assumption. In reply, it seems that the concept of "evidence against" is stretched if it includes simply an assertion that certainty is impossible.

Another argument is that the belief-function theory applies only when it is possible to obtain *all* the evidence, even, if necessary, the final diagnosis. If this reply is accepted, then belief functions are limited in their ability to represent second-order uncertainty. In particular, they cannot represent cases in which the evidence is truly limited.[4]

It is possible to define a set of functions that permit translation if we drop the condition that $U(1)>0$. In particular, let us assume that $Q(1)=B$ and $Q(P)=(1-B)/N$ for the N values of P such that $P<1$. B is thus the parameter that characterizes $Q(.)$. When B is 1, A (from equation 2) is 1, and when B is 0, A is 0 because there is no evidence for the proposition. (Also, $Q(1)=0$ when B is 0.) We seek a function relating A and B that preserves the translation between systems.

Note first that the denominator of equation 1, $\sum_P Q_1(P) \cdot Q_2(P)$, is $B_1 \cdot B_2 + N \cdot (1-B_1) \cdot (1-B_2)/N^2$. This follows from our definition of $Q(.)$. The first term is for $P=1$ and the next term is for the N points at which $P<1$. (The second term of course simplifies to $(1-B_1) \cdot (1-B_2)/N$.) Hence,

$$B=Q(1)=\frac{B_1 \cdot B_2}{B_1 \cdot B_2+(1-B_1) \cdot (1-B_2) \cdot N}.$$

Inverting, we get

$$1/B=1+\frac{(1-B_1) \cdot (1-B_2) \cdot N}{B_1 \cdot B_2},$$

or

$$\text{(3)} \qquad \frac{(1-B)}{B}=\frac{(1-B_1) \cdot (1-B_2)}{B_1 \cdot B_2/N}.$$

It is apparent that the translation will work only if $N=1$ and $(1-A)=f(B)=[(1-B)/B]^r$ for any nonzero r. However, if $N>1$, there is no rule that will work. Hence, the only form of $Q(.)$ that will satisfy the condition is one with all the probability concentrated at $P=1$ and at a single other value of P. ($P=0.5$ would make sense.) The same conclusion follows if we examine $Q(P)$ for some other value of P less than 1.

Suppose now that $Q(1)>0$. We now let $Q(1)=B+(1-B)/N$, and $Q(P)=(1-B)/N$ (as before) for the $(N-1)$ values of P such that $P<1$. Using equation 1 for some value of P less than 1, we get,

$$\frac{1-B}{N}=\frac{\frac{1-B_1}{N}\cdot\frac{1-B_2}{N}}{\left[B_1+\frac{1-B_1}{N}\right]\cdot\left[B_2+\frac{1-B_2}{N}\right]+\frac{1-B_1}{N}\cdot\frac{1-B_2}{N}}.$$

Inverting gives

$$\frac{N}{1-B}=(N-1)+\frac{\left[B_1+\frac{1-B_1}{N}\right]\cdot\left[B_2+\frac{1-B_2}{N}\right]}{\frac{1-B_1}{N}\cdot\frac{1-B_2}{N}}.$$

Only if $N=1$, this simplifies to

$$\frac{1}{1-B}=\frac{[(1-B_1)+B_1]\cdot[(1-B_2)+B_2]}{(1-B_1)\cdot(1-B)}=\frac{1}{1-B_1}\cdot\frac{1}{1-B_2}.$$

In this case, when $N=1$, the basic idea no longer makes sense; P would have to have a value of 1. Hence, there is no sensible form of $Q(.)$ that allows translation if $Q(1)>0$ and if the other values of $Q(.)$ are all equal.[5]

It does not seem hopeful to try to remedy the situation by permitting unequal values of $Q(P)$ for $P<1$. Repeated attempts have yielded neither a function than meets the required conditions nor a proof that no such function exists.

It is clear that belief functions are not a necessary consequence of the idea of second-order probability. Because second-order probabilities are consistent with Bayesian principles and apply to the same sorts of situations that belief functions apply to, it would appear that the theory of belief functions requires additional justification as a normative model. In comparison to the Bayesian system as a whole, belief functions appear to be *restrictive* in the constraints placed on the form of $Q(.)$. The restrictions they place on $Q(.)$ are not even acceptable within the spirit of the belief-function theory itself. It is possible that there is no possible representation of belief functions in terms of second-order probabilities.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Science Foundation (principal investigators: J. B. and John C. Hershey). The idea of second-order uncertainty as missing information was introduced to me by Debbie Frisch. Shimon Schocken and John C. Hershey provided useful criticisms and substantive suggestions.

NOTES

[1] Einhorn and Hogarth (1985) also note that extreme probabilities such as 0.001 or 0.999 tend to be treated as closer to 0.5 when they are felt to be ambiguous.

[2] We might have third-order probabilities, $R(Q(P))$, and so on. However, the use of second-order probabilities as an analytic tool does not require – logically or pragmatically – the use of higher-order probabilities. I doubt that third-order probabilities, etc., will be of much utility in decision-analysis or elsewhere.

[3] We deal here only with what Shafer calls simple support-functions. Translation between second-order probabilities and belief functions must be possible for simple support functions if it is possible at all.

[4] There are other problems with this proposal that will be discussed later.

[5] If it had been possible that $N=2$, we would have been able at least to represent second-order uncertainty in which the only evidence available was total evidence, so that P would either be 1 or 0 after it was obtained. We therefore see that this is no longer possible.

REFERENCES

Baron, J.: 1985, *Rationality and Intelligence*, Cambridge: Cambridge University Press.

Brown, R. V.: 1986, *Assessment Uncertainty and the Firmness of Information: A Decision-Oriented Methodology*, Falls Church, VA: Decision Sciences Consortium, Inc.

Einhorn, H. J. and Hogarth, R. M.: 1985, 'Ambiguity and Uncertainty in Probabilistic Inference', *Psychological Review* **92**, 433–461.

Ellsberg, D.: 1961, 'Risk, Ambiguity, and the Savage Axioms', *Quaterly Journal of Economics* **75**, 643–699.

Gärdenfors, P. and Sahlin, N.-E.: 1983, 'Decision Making with Unreliable Probabilities', *British Journal of Mathematical and Statistical Psychology* **36**, 240–251.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A.: 1971, *Foundations of Measurement* (Vol. 1), New York: Academic Press.

Lindley, D. V., Tversky, A., and Brown, R. V.: 1979, 'On the Reconciliation of Probability Assessments', *Journal of the Royal Statistical Association A*. **142**, 146–180 (with commentary).
Raiffa, H.: 1968, *Decision Analysis*, Reading, Mass.: Addison-Wesley.
Savage, L. J.: 1954, *The Foundations of Statistics*, New York: Wiley, 1954.
Shafer, G.: 1976, *A Mathematical Theory of Evidence*, Princeton, N. J.: Princeton University Press.
Shafer, G. and Tversky, A.: 1985, 'Languages and Designs for Probability Judgment', *Cognitive Science* **9**, 309–339.

Psychology Department,
University of Pennsylvania,
Philadelphia, PA 19104–6196,
U.S.A.