On Probability Intervals*

Judea Pearl
University of California, Los Angeles

ABSTRACT

The apparent failure of individual probabilistic expressions to distinguish between uncertainty and ignorance, and between certainty and confidence, have swayed researchers to seek alternative formalisms, where confidence measures are provided explicit notation. This paper summarizes how a causal networks formulation of probabilities facilitates the representation of confidence measures as an integral part of a knowledge system that does not require the use of higher order probabilities. We also examine whether Dempster-Shafer intervals represent confidence about probabilities.

KEYWORDS: ***probability, interval probability, Dempster-Shafer theory, causal networks***

INTRODUCTION

People make a distinction between sure and unsure probabilistic judgments. For example, everyone would agree that a typical coin has a 50% chance of turning up heads, while most people would hesitate to assign a definite probability to a coin produced in a gambler's basement. For that reason we sometimes feel more comfortable assigning a range rather than a point estimate of uncertainty, thus expressing our ignorance, doubt, or lack of confidence in the judgment required. We may say, for example, that the probability of the coin turning up heads lies somewhere between 60% and 40%, having no idea whether or how the coin was biased.

The apparent failure of individual probabilistic expressions to distinguish between uncertainty and ignorance, certainty and confidence, have swayed researchers to seek alternative formalisms, where confidence measures are provided explicit notation (Shafer [1]). In a recent paper (Pearl [2]), I have attempted to demonstrate how the causal networks formulation of probabilities facilitates the representation of confidence measures as an integral part of one's

* This work was supported in part by Grants DCR-8313875 and N00014-87-K-2029.

Address correspondence to Judea Pearl, Computer Science Department, UCLA, Los Angeles, California 90024.

52 Vanderbilt Ave., New York, NY 10017 0888-613X/88/$3.50

knowledge system, requiring no specialized notation or the use of higher-order probabilities (Kyburg, this volume).

In this note I will summarize my earlier argument (Pearl [2]) and then examine whether probability intervals in the Dempster-Shafer (D-S) formalism represent two components of confidence measures: insecurity due to uncertain contengencies and ignorance for lack of a complete model. My conclusion is that they do not. Although the presence of a nonzero D-S interval is sufficient to indicate some incompleteness in the probability model, the narrowing of the interval or even its disappearance does not indicate removal of ignorance or increased confidence in the model.

ON INSECURITY DUE TO UNCERTAIN CONTINGENCIES

The starting point is a claim that probabilistic statements such as $P(A) = p$ are in themselves empirical events of no less stature than other sentences reporting empirical observations. While not referring to an event open to full public scrutiny, these statements do, nevertheless, report outcomes of genuine experiments, namely, the mental procedures involved in assessing the belief of a given proposition A. Thus, stating "event A has a chance p of occurring" is equivalent to stating "the mental event of computing the likelihood of A has produced the outcome p."

Having endowed probabilistic statements with event status neutralizes the syntactic objection against writing sentences such as $P[P(A) = p]$. Both the square brackets and parentheses enclose arguments of the same type—empirical events. True, the latter event is external, while the former is personal. However, this distinction is not a barrier to useful semantics; the adoption of a computational model of knowledge representation (e.g., semantic networks, causal models) permits us to specify the mental procedures involved in belief assessments with the same clarity and precision with which we specify experimental procedures in a laboratory setting. What remains to be done is, first, to explain what renders the event $P(A) = p$ an unknown, random event, rather than a fixed outcome of a stable procedure and, second, to explicate more precisely the mental procedures involved in making the two assessments, $P(A)$ and $P[P(A) = p]$.

A paradigm answering the first question has been suggested by de Finetti [3] and has been guiding the Bayesian interpretation of confidence measures for over a decade (Spiegelhalter [4], Heckerman and Jimison [5]). The basic idea is that the event $P(A) = p$ is perceived to be a random variable whenever the assessment of $P(A)$ depends substantially on the occurrence or nonoccurrence of some other events modeled by the system called *contingencies.* In the words of de Finetti:

> The information apt to modify the probability assessed for an event E—

> in so far as the observation of H_i makes us change from $P(E)$ to $P(E|H_i)$—can make us view the H_i's as sort of "noisy" signals concerning the occurrence and non-occurrence of the event E.

Adopting this interpretation, one can show that the procedure involved in the assessment of $P[P(A) = p]$ is no different from that involved in the assessment of $P(A)$ and, moreover, that the very information used for calculating $P(A)$ is sufficient for calculating the confidence interval associated with the statement $P(A) = p$. This is based on the observation that by specifying a causal model for predicting the outcome A, we automatically specify the variance of that prediction. Formally, if C is a set of contingencies affecting A, then knowing $P(A \mid c)$ and $P(c)$ permits us to simulate the behavior of $P(A \mid c)$ as C takes on various realizations c with their associated probabilities $P(c)$. The histogram of $P(A \mid c)$ then defines the variance of $P(A)$.

In other words, when a person encodes probabilistic knowledge as a causal model of interacting variables, that person automatically specifies, not merely the marginal and joint distributions of the variables in the system, but also a set of future scenarios, describing how these probabilities would vary in response to future eventualities. It is this implicitly encoded dynamics that renders probabilistic statements random events, admitting distributions, intervals and other confidence measures. Thus, the notions of insecurity and doubt are intrinsic and indigenous to classical probabilistic formulation; no second-order probabilities or specialized notational machinery are required to reinstate them where they flourish so naturally.

ON THE DEMPSTER-SHAFER INTERVALS

I have argued (Pearl [2]) that the Dempster-Shafer (D-C) interval does not represent the degree of insecurity people feel toward the assessment of point probability values. My argument was that people's insecurity is often associated with a high degree of sensitivity to unknown contingencies, and that such sensitivity is describable in traditional models of probability theory and, since the D-S interval vanishes whenever we are in possession of a complete probabilistic model, it could not possibly reflect this component of people's insecurity.

Reiterating the example given in my earlier paper (Pearl, [2]), suppose we know that a given coin was produced by a defective machine—that precisely 49% of its output consists of double-headed coins, 49% are double-tailed coins, and the rest are fair. This description constitutes a complete probabilistic model that predicts that the outcome of the next toss will be heads with probability 50% and alerts us to the fact that the prediction is extremely susceptible to new information regarding the nature of the coin. Most people will hesitate to commit a point estimate of 50% to the next outcome of the coin, as is attested by the natural tendency to lower one's bet, on heads or tails alike, compared with bets wagered on a fair coin. Most people would rather wait for some clue or toy with introspective analysis reflecting on the coin type. The D-S theory nevertheless assigns the next outcome a belief of 50%, with zero belief interval. Now imagine that we toss the coin twice and observe a tail and a head. This immediately implies that the coin is fair, and hence most people would regain confidence to assign the next toss a 50% chance of turning up heads. Yet such narrowing of the confidence interval would remain unnoticed in the D-S formalism; the theory will again assign the next outcome a belief of 50% with zero belief interval.

Next we examine the notion of *ignorance* due to model incompleteness. The D-S interval is often interpreted to portray the degree of ignorance we have about probabilities—namely, the degree to which the information we lack prevents us from constructing a complete probabilistic model of the domain. If this were so, then the D-S approach would indeed have a definite advantage over Bayesian methods, which always provide point probabilities. Unlike the latter, which often give one a false sense of security in the model, the D-S interval would have served as a warning device, distinguishing beliefs based on well-founded probabilities from those based on partially specified models.

Unfortunately, the D-S intervals have little to do with ignorance, nor do they represent *bounds* on the probabilities that would ensue once ignorance is removed. This can be demonstrated using the classical three-prisoners puzzle.[1]

The story involves three prisoners A, B, and C awaiting their verdict, knowing that one of them will be found guilty and the other two released. Prisoner A asks the jailer, who knows the verdict, to pass a letter to some other prisoner who is to be released. Later, prisoner A asks the jailer for the name of the letter recipient and, having learned that the jailer gave the letter to prisoner B, the problem is to assess the belief that A is the one found guilty.

The problem can be formulated in terms of three mutually exclusive and exhaustive propositions G_A, G_B, and G_C, where G_i stands for "prisoner i was found guilty." Coupled with these, we also have the jailer's testimony, which could have been either "B" or "C" and so can be treated as a bivalued variable L (connoting "letter recipient") taking on the values $\{L_B, L_C\}$.

In the Bayesian treatment of the problem, one assumes equal prior probabilities on the component of G, $\pi(G_A) = \pi(G_B) = \pi(G_C) = 1/3$, and $P(L_B \mid G_A) = 1/2$; namely, if A were guilty, the jailer would choose the letter recipient at random, giving an equal chance to B and C. These two assumptions yield the answer $P(G_A \mid L_B) = 1/3$, meaning that the jailer's testimony is totally irrelevant relative to A's prospects of being released. If, on the other hand, the letter is not handed at random but the jailer prefers B (or C), then the posterior

[1] The following analysis contains excerpts from my forthcoming book *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufman, 1988).

probability $P(G_A | L_B)$ would vary from zero (if B is avoided) to ½ (if C is avoided).

In the D-S treatment of the problem we do not assume values for the prior or conditional probabilities unless we have evidence to substantiate these values. For example, if we have good reason to believe that the testimonies in the trial are equally supportive of either prisoner's innocence, then, and only then, we would take the liberty of assigning equal weights to the components of G, $m(G_A) = m(G_B) = m(G_C) = 1/3$. Assuming this is the case in our story, still, having no idea of the process by which the letter recipient was selected prevents us from completing the model and yields Bel (G_A) = Bel $(\neg G_A) = 1/2$; reflecting a zero interval yet an answer different from that of Bayesian analysis.

This disparity is not surprising in view of the fact that we have an incomplete probabilistic model on our hands, as the process by which B was selected remains unspecified. Conservatively speaking, it is quite possible that the jailer's choice was not random but marred by a deliberate attempt to avoid choosing C whenever possible. Under such extreme circumstances, the jailer's answer L_B could only be avoided one-third of the time (when B is guilty), thus leaving A and C an equal chance of being the condemned. What may sound somewhat counterintuitive is that, from among all possible ways of completing the model, D-S theory appears to select this extreme and unlikely model, which also happens to be the one that puzzle books repeatedly warn us to avoid.

Actually the D-S theory never attempts to complete the model, and, although the jailer's testimony causes all the weight to be committed to singleton hypotheses, $m(G_A) = m(G_C) = 1/2$, the model remains only partially specified, as we still are in ignorance regarding the letter delivery process. Knowing the selection process is important because, in Bayesian analysis, it could sway the posterior probability $P(G_A | L_B)$ all the way from zero to ½. Yet the interval $Pl(G_A)$ − Bel (G_A) is zero, giving one the false impression that the answer Bel $(G_A) = 1/2$ is based on a complete model (with the jailer attempting to avoid C whenever possible).

The disparity between the answers produced by the two formalisms stems not from the weight distribution but rather from the semantics of these answers. While the probabilistic approach interprets "belief in A" to mean the conditional probability that A is *true* given the evidence e, the D-S approach calculates the probability that proposition A becomes *provable* given the evidence e. Phrased another way, it computes the probability that some set of hypotheses suggested by the evidence would materialize (e.g., that the judges become convinced by an alibi), from which the truth of A can be derived out of logical necessity. Thus, instead of the conditional probability $P(A | e)$, the D-S theory computes the probability of the logical entailment $e \models A$. The two could be made as far apart as one wishes, depending on the choice of compatibility relationships by which proofs are constructed.

Thus, the disappearance of the D-S interval $Pl(A)$ − Bel (A) does not mean the removal of ignorance. It simply means that, based on the logical abstraction chosen to represent compatibility relationships, the available evidence could not simultaneously be compatible with A and its negation $\neg A$. It is curious to note that applying the same interpretation to Bayesian models yields an interval that *never* vanishes, because, barring extreme probabilities, a body of (noisy) evidence is always combatible with both a proposition and its negation.

CONCLUSION

In conclusion, D-S intervals might have a place in the analysis of evidence, but they do not possess all the qualities that the literature often wishes them to have. In particular, they do not represent insecurity about probability assessments or ignorance about missing information. The former can be obtained from the traditional representation of Bayesian networks, while the latter can be obtained from the bounds produced by Nilsson's probabilistic logic (Nilsson [6]).

References

1. Shafer, G., *Mathematical Theory of Evidence,* Princeton University Press, Princeton, N.J., 1976.
2. Pearl, J., Do we need higher-order probabilities and, if so, what do they mean?, *Proceedings, AAAI Workshop on Uncertainty in AI,* Seattle, Wash., July 1987, pp. 170–179.
3. De Finetti, B., Probabilities of probabilities: a real problem or a misunderstanding, in *New Developments in the Applications of Bayesian Methods,* (A. Aykac and C. Brumet, Eds.), North-Holland, Amsterdam, 1977, pp. 1-10.
4. Spiegelhalter, D. J., Probabilistic reasoning in predictive expert systems, in *Uncertainty in Artificial Intelligence* (L. N. Kanal and J. Lemmer, Eds.), North-Holland, Amsterdam, 1986, pp. 47-68.
5. Heckerman, D., and Jimison H., A perspective on confidence and its use in focusing attention during knowledge acquisition, *Proc. AAAI Workshop on Uncertainty in AI,* Seattle, Wash., 1987.
6. Nilsson, N., Probabilistic logic, *AI,* **28**(1), 71-87, 1986.