In *Annals of mathematics in Artificial Intelligence,* Special Issue on
Representations on Uncertainty, Vol. 32, 171-177, 2001.

TECHNICAL REPORT
R-279
February 2001

On Two Pseudo-Paradoxes in Bayesian Analysis

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

Introduction

This note discusses two problems that might be considered weaknesses of Bayesian analysis. The first was noted in a recent paper of Cozman (2000), and concerns the notion of "relevance" when complete probabilistic model is not available. The second, stimulated by an example due to Philippe Smets, concerns the interpretation of evidential reports that are cast as betting odds. I will describe the two problems, offer their resolution, and then argue that Bayesian analysis should retain its status as a powerful model of human reasoning.

1 Informational relevance in partially specified models

In standard Bayesian analysis, the notion of informational "relevance" is captured by the construct of probabilistic "dependence" (or conditional dependence), and is fully characterized when a complete probability model is available. We say that an event A is *relevant* to B when the discovery of A changes our belief in B, that is, $P(B|A) \neq P(B)$. Similarly, the notion of "irrelevance" is formalized using (conditional) "independence" (Dawid 1979)—a notion governed by a set of qualitative axioms called "graphoids" (Pearl 1988), which have intuitive appeal beyond the boundaries of probability analysis.

The apparent paradox I wish to discuss in this note surfaces when we do not possess a complete probability function. In such a state of ignorance (about probabilities), theory predicts that some of the graphoid axioms are violated, yet these violations are not reflected in actual reasoning—human judgment continues to conform to the dictates of the graphoid axioms.

Consider two jointly distributed (discrete) variables, X and Y, for which we have the conditional probabilities $P(y|x)$ for all x and y, but we lack any knowledge of the prior probability $P(x)$. With this state of partial ignorance, and assuming $P(y|x) \neq P(y|x')$ for some $x' \neq x$, X is judged to be relevant to Y, because the measurement $X = x$ would induce a different belief in $Y = y$ than the measurement $X = x'$. At the same time, measurement of Y leaves our uncertainly about X unaltered, for the bounds on $P(x|y)$ remain $[0, 1]$ for all findings $Y = y$. Thus, it appears that the axiom of symmetry, that X is relevant to Y if and only if Y is relevant to X, should be violated in this state of knowledge.[1] Yet such violation is not reflected in actual judgment.

Let $X = x$ stand for the existence of a technologically advanced civilization in some remote galaxy. Let $Y = y$ stand for receiving an intelligible radio signal from deep space. Assume, as is common among humanoids, that we have difficulty assigning a prior to X, though we can assess the conditional probability $P(y|x)$. If we associate our epistemic state of belief with the *set of probabilities* that are considered possible,[2] and if we equate irrelevance with the invariance of that set of probabilities, then we should classify Y as irrelevant to X; our uncertainty concerning x will remain unaltered by any observation of Y, since the bounds on $P(x)$ remain $[0, 1]$ before and after observing $Y = y$. But this is not supported in commonsense reasoning—receiving intelligible radio signals from deep space would no doubt evoke fancy speculations regarding the existence of intelligent civilization out there,[3] hence it would definitely be considered "relevant" to X.

This pattern of judgment prevails throughout science, where scientific theories are often conceived and formulated as a result of unexpected empirical findings. Such theories are not contemplated in advance, and definitely are not assigned prior probabilities. Still, it would be odd to say that, in the pre-discovery state of mind, obtaining those empirical findings would be judged irrelevant to the truth of the theory.

What accounts for this discrepancy? Is it that Bayesian analysis fails to represent prevailing pattern of scientific inference, or that the representation of partial knowledge in the form of probability intervals is incomplete? I will argue for the latter alternative.

Whereas representing a state of ignorance in the form of a set of probabilities is an effective construct for many purposes (see (Walley 1991; Cozman 1997)), one should not equate an agent's epistemic state with the boundaries of that set. The dynamics of each member of that set plays an important role in what the agent knows and believes, not less important than the dynamics of the boundaries.

To cast this assertion in concrete terms, let us return to the radio signal example and examine the dynamics of points in the interval $0 \le P(x) \le 1$. While the boundaries of this interval are not altered by observing $Y = y$, since $0 \le P(x|y) \le 1$, each interior point in this interval moves from $P(x)$ to

$$P(x|y) = \frac{LP(x)}{1 + (L-1)P(x)}$$

where L is the likelihood ratio $P(y|x)/P(y|x')$. Thus, assuming $L > 1$, every interior point undergoes some movement toward $P(x) = 1$, with the size of the movement varying from point to point. I suggest that it is this internal motion which accounts for our perception of Y as relevant to X, in this example. Ignoring this internal motion would amount to a loss of

[1] Cozman (2000) discusses additional violations of the graphoid axioms.

[2] Levi (1980) called such sets *credal sets*, and Walley (1991, chapter 9) called an event Y *epistemic irrelevant* if knowledge of Y does not change the agent's credal set, arguing that such knowledge would not affect the practical consequences of the agent's beliefs (see Cozman, 2000).

[3] Assuming, for argument sake, that $P(y|not\text{-}x) \neq 0$, say the possibility exists that a smart hacker may fabricate such signals.

valuable information, not the least of which is our judgment of relevance relationships, which is central for controlling attention and for choosing among alternative information sources.

At the same time, in order to envision this internal motion, an agent must first postulate one or several interior points in the set of possible probability functions, then examine their movement under information update. Thus, we see that there is some subtle wisdom to the sloppy advice often offered by Bayesians: "just assume any conventient prior." True, assuming any conventient prior may not lead to a valid posterior, but it helps uncover nevertheless the dynamics of belief updating that interval representations tend to mask.

2 Two Interpretations of evidential reports

> Suppose someone was murdered and you try to assess which of 4 suspects, denoted a, b, c, and d, is the actual killer, given that one and only one of the suspects committed the murder. Initially, you have no reason to suspect any one of the four more than another. Two pieces of evidence arrive:
>
> Evidence 1, denoted e_1, makes you believe strongly that the killer is one of a, b, or c, but it might still be d. In fact, looking at this evidence, you are willing to bet 2:1 that it is not d.
>
> Evidence 2, denoted e_2, makes you sure that the killer is one of a and d.
>
> What is your belief about the actual killer? Which of a or d you feel is the more likely killer?

This basic story,[4] in various shades and colors, has been devised to demonstrate that Bayesian analysis clashes with intuition and commonsense. Most people feel that a is more likely to be the killer, while Bayesian analysis (so the the argument goes) predicts that d is more likely.

Here is what makes most people conclude the a is more suspect than d. From the phrase "...e_1, makes you believe strongly...," people infer that e_1 made the class $\{a, b, c\}$ more suspect than it was before. Further, our willingness to bet 2:1 against d seems to imply that e_1 (partially) exonerates d from suspicion. Evidence e_2 merely removes b and c from consideration, but does not alter the relative degree of suspicion between a and d. The net result is that a's guilt is positively supported by the evidence, while d's guilt receives negative support.

To make this argument more compelling and concrete, we can imagine that $\{a, b, c\}$ are smokers, that d is a nonsmoker, and that e_1 is some indication that the killer smoked. We can equally imagine that e_2 arrives first, and e_1 second. e_2 removes b and c from consideration (say they had a strong alibi) and leaves a and d as the only suspects, with no reason to prefer one over the other. Now comes the evidence about smoking, e_1,—surely d should be deemed less likely to be the killer.

[4] The story was presented to me by Phillipe Smets, (November, 1999) in an e-mail message entitled "The danger of equiprobable priors." I took the liberty of making minor changes in the original text, so as to render the accompanying arguments more plausible. I am indebted to Phillipe for the example and for subsequent e-mail discussions, though he does not agree with my analysis (and conclusions).

How does Bayesian analysis treat this story? According to its critics, the analysis should proceed as follows. The input "you are willing to bet 2:1 that it is not *d*" should be translated into a statement about posterior probabilities, conditioned on e_1,

$$P(\{a,b,c\}|e_1) = 2P(\{d\}|e_1), \tag{1}$$

from which we readily obtain

$$P(\{a,b,c\}|e_1) = \frac{2}{3}, \quad P(\{d\}|e_1) = \frac{1}{3} \tag{2}$$

If we further assume that a, b, c are equally suspect, each will get a probability $1/3 \cdot 2/3 = 2/9$ of being the killer. The arrival of e_2 removes b and c from consideration but leaves the ratio $P(\{d\}|e_1) : P(\{a\}|e_1)$ the same (as in standard Bayes conditionalization), which yields

$$P(\{a\}|e_1, e_2) : P(\{d\}|e_1, e_2) = \frac{2}{9} : \frac{1}{3} = 2 : 3. \tag{3}$$

Thus, we finally see that, contrary to commonsense, d is deemed more likely to be the killer than a (by a 3:2 ratio).

Can this clash be reconciled?[5] It surely can.

Bayes' analysis remains faithful to commonsense, but commonsense is vulnerable to ambiguity in the interpretations of the key statement:

> "...looking at this evidence, you are willing to bet 2:1 that it is not *d*"

The interpretation expounded in the intuitive argument above, according to which e_1 supports *d*'s innocence, does not interpret the betting ratio literally, but assigns to it a likelihood-ratio interpretation:

$$P(e_1|\{a,b,c\}) : P(e_1|\{d\}) = 2 : 1 \tag{4}$$

which renders a a more likely suspect:

$$P(\{a\}|e_1) = 2P(\{d\}|e_1)$$

This interpretation is compelled in fact when we consider e_1 as indicative of the killer's smoking habits. If our preference toward $\{a,b,c\}$ originates from comparing the smoking behavior of the killer to those of the suspects, then e_1 should be judged twice more likely to have been produced by *any one* of a, b, and c, than by d, namely,

$$P(e_1|\{a\}) = P(e_1|\{b\}) = P(e_1|\{c\}) = 2P(e_1|\{d\})$$

giving (assuming equal prior probabilities)

$$P(\{a\}|e_1) = P(\{b\}|e_1) = P(\{c\}|e_1) = 2P(\{d\}|e_1),$$

[5] I am somewhat embarrassed to offer a resolution that should be obvious to most practicing Bayesians and to any reader of my book (Pearl 1988, Sections 2.2.2 and 2.3.3). However, considering that pseudo-paradoxes of this type tend to reappear periodically in the literature, a reaffirmation of fundamentals might be in order.

and the posterior probability of d should calculate to

$$P(\{d\}|e_1) = \frac{1}{9}, \tag{5}$$

not to 1/3 as stated in (2). This interpretation is incompatible with the posterior probability interpretation of (1), because, according to (5), the odds of betting against d should have been 8:1 and not 2:1 as stated in the story. Moreover, taking the input statement literally, betting 2:1 against d implies that d's guilt received positive, not negative support, because the belief in d's guilt went up from 1/4 to 1/3.

In summary, we see that the clash that emerges from this story is not a clash between Bayesian vs. intuitive reasoning but, rather, between two legitimate interpretations of the input sentence:

> "... looking at the evidence, you are willing to bet 2:1 that it is not d"

The first interprets this sentence literally, as quantifying the $final$ beliefs resulting from e_1, expressed in terms of one's willingness to bet on the propositions $\{d\}$ vs. $\{a, b, c\}$. The second interprets this sentence colloquially, as describing the incremental *change* in one's beliefs and, hence, as quantifying the relative strength of evidential support that $\{d\}$ and each of its rival alternatives receives from e_1, and from e_1 alone.

If we take the literal interpretation, then Bayesian analysis encodes the input statement as a ratio between two posterior probabilities (as in (1)), and helps us derive the logical implication of this interpretation, i.e., d is more likely to be the killer than a (as in (3)). If, however, we take the colloquial interpretation, then Bayesian analysis is again at our service; it encodes the input sentence in the form of a likelihood ratio (as in (4)), and again derives the correct implication, i.e., a is more likely to be the killer than d.

The tension between these two interpretations dates back to Jeffrey's conditionalization (Jeffrey 1965) which was devised to handle belief updating based on non propositional observations (see Pearl (1988, pp. 62–69)), that is, observations summarized in terms of the final probabilities they induce. Goldszmidt and Pearl (1996) analyzed the semantics of these interpretations and distinguished between two types of evidential reports, Type-J (connoting "Jeffrey") and Type-L (connoting "likelihood ratio"). Type-J, which corresponds to our literal interpretation in terms of final beliefs, requires considerations of the agent's pre-observation beliefs, and was characterized therefore as "All things considered". Type-L, which corresponds to our colloquial interpretation in terms of belief changes, is based exclusively on the observation at hand, and was characterized as "Nothing else considered". Pearl (1988, pp. 44–47) discusses the tendency of people to quantify incremental changes of belief in terms of absolute probabilities, and how the reported probabilities should be converted back to likelihood ratios.

This still behooves us to explain why it is so easy to lure readers into the colloquial interpretation in terms of evidential support, or belief change, and ignore the literal interpretation in terms of final beliefs in a betting situation. After all, the story states explicitly that the 2:1 ratio stands for one's willingness to bet, and makes no mention at all of strength of evidence or belief change. The answer lies in what may seem to be an innocence, inconsequential phrase: "... looking at the evidence, you will be willing to bet ...". Can we look at the evidence e_1 and decide how to bet on d versus not-d? Can we examine a laboratory report concerning traces of tobacco on the victim's body and decide, considering that d does not smoke, how to bet on d versus not-d? The answer is: "No!"; because such decision must take into account other factors beside the report, for example, how many smokers and non-smokers are suspect, whether other suspects have solid alibi, etc. Forensic experts are undoubtedly instructed to ignore such considerations when judging how likely the killer is to be a smoker. Therefore when we tell someone: "Looking at this evidence," the listener expects to find a summary of that evidence and that evidence alone. Finding a statement about betting behavior, the listener has no choice but to translate this statement into a Type-L evidential report, namely, a likelihood ratio.

The likelihood ratio is the only interpretation that rests strictly on the relationships between the laboratory findings and the smoking behavior of the killer, one that is not contaminated with previous beliefs about suspect d or his companions.[6] To perform this translation, the listener constructs a hypothetical, standard betting situation in which one starts with a neutral (50–50) position on whether the killer is a smoker or non-smoker and, upon seeing the evidence e_1, one ends up with the specified betting odds: 2:1 in favor of a smoker. It is this hypothetical interpretation of the input betting odds that leads listeners to the presumption that e_1 supports d's innocence (as reflected in (4)) rather than d's guilt (as implied by the literal interpretation of (1).)

3 SUMMARY

We have examined two aspects of Bayes's analysis that, at first glance, seem incompatible with human reasoning. The first concerned the practice of postulating prior probabilities in cases of complete ignorance (about probabilities). We showed that this practice may serve a useful cognitive function—the detection of informational relevance among potentially observable events. Such relevance relation may remain undetected in the interval representation of ignorance, where sets of probabilities are used to encode agents beliefs. The second aspect concerned the proper handling of evidential reports that are cast as betting odds. We showed that often cited paradoxes in such cases have little to do with the assumption of equiprobable priors. Rather, they reflect a clash between two legitimate interpretations of evidential reports that must be decided at the onset of any analysis. We also showed that Bayes's analysis is well equipped to handle both interpretations, and that the likelihood ratio interpretation is the more natural one of the two.

References

[Cozman, 1997] F.G. Cozman. Robustness analysis of bayesian networks with local convex sets of distributions. In D. Geiger and P. Shenoy, editors, *CUAI*, pages 108–115. Morgan

[6] The likelihood ratio enjoys these stable features partly because it invokes conditional probabilities along the direction of causal influences. (In our example, the probability of obtaining the laboratory finding, conditioned on the killer's smoking habits.) The causal slant of the likelihood ratio is rarely acknowledged in the statistical literature, though one can hardly find a likelihood function in which the conditioning events are the effects, rather the causes. The general stability of causal relationships is discussed at length in Pearl (2000, pp. 24–25).

Kaufmann, San Francisco, CA, 1997.

[Cozman, 2000] F.G. Cozman. Separation of sets of probability measures. In *Proceedings of the Sixtheenth Conference on Uncertainty in Artificial Intelligence*, pages 107–114. Morgan Kaufmann, San Francisco, CA, 2000.

[Dawid, 1979] A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B*, 41(1):1–31, 1979.

[Goldszmidt and Pearl, 1996] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1–2):57–112, July 1996.

[Jeffrey, 1965] R. Jeffrey. *The Logic of Decision.* McGraw-Hill, New York, 1965.

[Levi, 1980] I. Levi. *The Enterprise of Knowledge.* MIT Press, Cambridge, Massachusetts, 1980.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York, 2000.

[Walley, 1991] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.