PART VII

PROBABILITY AND CAUSALITY

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

Probabilistic Reasoning in Expert Systems Reconstructed in Probability Semantics

Roger M. Cooke

Delft University of Technology

Probabilistic reasoning is traditionally represented by inferences of the following form (also called probabilistic explanations):

1) $$\frac{\begin{array}{l} P(A \mid B) = q \\ B(j) \end{array}}{A(j)}$$

where A and B are one-place predicates in a first order language, P(A | B) is the conditional probability of observing A among individuals having property B, and q is close to one.

This argument is not logically valid, as the premises may be true while the conclusion is false. Moreover, as it stands, the premises do not even make the conclusion plausible. It may be the case that 90% of the population dies before the age of 85, and that individual j is a member of the population, but this in itself does not make it plausible that individual j will die before the age of 85. There may be all sorts of other statistically relevant information which decisively influence j's probability of dying before the age of 85.

The problem of representing valid probabilistic reasoning is traditionally tied up with the problem of the reference class: to which population should j be assigned in order to determine the probability (in the sense of limiting relative frequency) that j has the property A? For example, we know that age, sex and smoking habits are statistically relevant for longevity. If j is a 40 year old male non-smoker, can we determine "the" probability of j dying before 85 by observing the relative frequency of death before 85 in the population of male non-smokers of age 40?

If we have some method to determine "the best" reference class for "estimating" P(A(j)), then we may also use this method to identify valid probabilistic inferences. We accept as valid schemes of the form (1) only when property B defines the "best" reference class, and when q is close to one. In this way we ensure that the conclusion has high probability. The same method could be used to construct "probabilistic

PSA 1986, Volume 1, pp. 409-421

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

inference engines" to drive expert systems. Indeed, in trying to capture the reasoning of an expert in a specific domain, the designer of an expert system is immediately confronted with the problem of representing valid probabilistic inference.

Hempel's requirement of maximal specificity (1968) is representative of attempts to define valid probabilistic reasoning via a solution to the problem of the reference class. His solution involves constraints on the state of knowledge at the time an inference is made. Roughly, the requirement of maximal specificity says that $P(A \mid B)$ must equal $p(A \mid B')$ whenever B' is a "maximally specific statistical predicate" such that $B'(j)$ holds.

This type of solution is unsatisfactory for three reasons.
a) A valid probabilistic inference of the form (1) may become invalid when new statistical knowledge is acquired. The new knowledge may even involve frequencies in reference classes to which j does not belong (Cooke 1981).
b) Valid probabilistic inferences should be valid theorems in probability, regardless of how probability is interpreted. If (1) were valid in Hempel's sense, then it is certainly not a theorem that $P(A(j)) = q$; indeed, it is not even clear how $P(A(j))$ should be defined. In a frequency interpretation, $P(A(j))$ is either undefined, or defined through a valid argument of the form (1); albeit that validity may not be preserved under conservative increases of knowledge. If probability is interpreted subjectively, then $P(A(j))$ can be defined in a variety of ways, but it need have no relation to the number q in a valid inference of the form (1).
c) Probabilistic reasoning in expert systems provides us with inferences which cannot be analysed in terms of the scheme (1). These inferences involve not only a "probability conditional" as a major premise but the minor premise may be probabilistic as well. In other words, instead of the premise $B(j)$ we might have the premise "$P(B(j)) = r$".

The thrust of this paper is to de-couple the problem of valid probabilistic inference from the problem of the reference class, and to interpret valid probabilistic inference as valid theorems in probability. The probability semantics of Łos (1963) is used for this purpose. The inference schemes in this solution differ somewhat from (1). Probability will be assigned to sentences as well as one-place predicates. Moreover, several "probability conditionals" can be distinguished in Łos' framework, and it turns out that the ordinary conditional probability statements for one-place predicates are just not the right conditionals for probabilistic inference. With this more powerful formalism it is also possible to analyse certain aspects of probabilistic reasoning in expert systems.

Section 1 reviews Łos' probability semantics in a suitably simplified form. Section 2 applies this to the problem of reconstructing probabilistic inference. Section 3 reviews the relevant aspects of probabilistic reasoning in the expert system MYCIN, and Section 4 uses the foregoing analysis to reconstruct this reasoning in the framework of probability semantics. A final section offers conclusions.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

1. Łos' Probability Semantics

In this section we give a brief account of Łos' probability semantics. For our purposes it suffices to restrict ourselves to very simple first order languages and to one-place predicates.

A first order language L is said to be *of finite domain with cardinality D* if, in addition to the usual axioms of the first order predicate logic, L contains axioms requiring the existence of exactly D distinct individuals each having distinct names in L. Let M(L) denote the set of models of L, (we assume that M(L) is a set), $\wp(M(L))$ the powerset of M(L). Let S(L) denote the set of sentences of L, and let A be a one-place predicate in L. Given a model $M \in M(L)$, we define:

$$P_M(A) := |A^M|/D,$$

where $|A^M|$ is the cardinality of the interpretation A^M of A in M. $P_M(A)$ is the relative frequency with which we would observe the property A in M in a long sequence of random samples (with replacement) from the domain of M.

Let P: $\wp(M(L)) \rightarrow [0,1]$ be a probability measure on $\wp(M(L))$. For $Q \in S(L)$ we define:

$$P(Q) := P\{M \in M(L) \mid M \models Q\}$$

For any one-place predicate A in L, we define:

$$P(A) := \sum_{M \in M(L)} P_M(A)P(M)$$

We shall call P a *probability model for L*. It is natural to think of P_M as an objective relative frequency within M, and to think of P as a subjective probability over the set M(L). (Of course, we could replace P_M with an arbitrary probability over the subsets of the domain of M, however, the above is natural for the applications in Section 4.) The conditional probability $P(A \mid B)$ is evaluated by the quotient P(A and B)/P(B). Note that

$$P(A \mid B) \neq \sum_{M \in M(L)} P_M(A \mid B)P(M).$$

If $B^M = 0$, for some M with positive probability, then the right hand side will not be meaningful, whereas the left hand side may well be meaningful. If Q, $R \in S(L)$, then of course $P(Q \mid R) = P(Q \text{ and } R)/P(R)$.

2. Probabilistic Reasoning

From now on we assume that A and B are one-place predicates in a first order language L of finite domain with cardinality D. Łos's probability semantics enable us to isolate at least four distinct semantic probability conditionals. Consider the following statements in the metalanguage:

i) $P(A \mid B) = q_1$
ii) $Vx\ P(A(x) \text{ and } B(x)) = q_2P(B(x))$
iii) $P(Vx(B(x) \rightarrow A(x))) = q_3$
iv) $Vx\ P(B(x) \rightarrow A(x)) = q_4$

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

In (ii) and (iv) the operator "Vx" does not range over elements in a domain, but rather ranges over the individual names in L. The restrictions on L are intended to make this type of quantification a useful operation.

All of these statements are semantical, that is, they are not sentences in the formal language L, but are statements about a particular probability model P for L. Nonetheless, any of them could in principle be used in probabilistic reasoning by someone whose belief state corresponded with P. Since conditional probability is not an operation on sentences in the formal language (conditionalization cannot be nested, whereas the material implication can be), it is hardly surprising that probabilistic reasoning is inherently "semantical".

(ii) may not look like a probabilistic conditional at first sight, but it is equivalent to the statement "for all names x, if $P(B(x)) \neq 0$, then $P(A(x) \mid B(x)) = q_2$".

Assuming that (i) - (iv) hold, the reader can easily verify the following trivial relations. If one of the q's = 1, then they all equal one. $q_1 = 0$ is equivalent with (only) $q_2 = 0$. If $q_4 = 0$, then in every model with positive probability, every element has the property B and no element has the property A. (ii) may hold for some q while (iv) does not hold for any q . Indeed, let j be such that $P(B(j)) = 0$. Then $P(B(j) \rightarrow A(j)) = 1$, so (iv) can hold only if (iii) holds with $q_3 = 1$. However, (ii) may hold when $q_3 \neq 1$ and $P(B(j)) = 0$.

Since if $P(B(j)) = 0$, for some j, (iv) can only hold with $q_4 = 1$, (iv) is not very interesting for probabilistic reasoning. (iii) is also not very interesting, as B can "almost" be a subset of A in every model while (iii) holds with $q_3 = 0$. (i) and (ii) are the only interesting candidates for the role of major premise in probabilistic arguments.

It is easy to verify that (i) may hold while (ii) fails for every q_2. However, if (ii) holds, then the following Fubini theorem shows that $q_2 = P(A \mid B)$, if this latter quantity is defined. We introduce the following notation:

$P(A \| B) = (\geq)\ q$ if for all names x $P(A(x)$ and $B(x)) = (\geq)\ qP(B(x))$,
and $P(B) \neq 0$.

<u>Theorem 1</u>: If $P(A \| B) = (\geq)\ q$, then $P(A \mid B) = (\geq)\ q$.

<u>Proof</u>: Put $X_{B,j}(M) = 1$ if $M = B(j)$ and $= 0$ otherwise. With abuse of notation, let D denote the set of names of L and the cardinality of this set.

$$\begin{aligned}\Sigma_D P(B(j)) &= \Sigma_D \Sigma_{M(L)} X_{B,j}(M)P(M) = \Sigma_{M(L)} \Sigma_D X_{B,j}(M)P(M)\\ &= \Sigma_{M(L)} |B^M| P(M) = \Sigma_{M(L)}\ P_M(B)P(M)D\\ &= P(B)D.\end{aligned}$$

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

The definition of P(A || B) implies

$$q\sum_D P(B(j)) =(\leq) \sum_D P(A(j) \text{ and } B(j)).$$

Using the above, the theorem now follows. ///

Note that finiteness of D is used in reversing the order of summation in the above proof. If D is not finite a uniformity condition would be required. Note also that the theorem would not hold if P_M were an arbitrary probability measure on the domain of M.

Remark: If P(A || B) =(≥) q and P(B(j)) =(≥) r, then P(A(j)) ≥ qr. If equality holds in the conditions and if for all M with positive probability $A^M \quad B^M$, then P(A(j)) = qr.

Proof: P(A(j)) ≥ P(A(j)) and B(j)) = P(A(j) | B(j))P(B(j)), with equality holding if $A^M \quad B^M$ for all M with positive probability.

Theorem 1 shows that P(A || B) ≥ q is a special case of P(A | B) ≥ q. Moreover, P(A || B) = q forces the predicate B to behave a bit like a maximally specific predicate in Hempel's sense. Indeed, if P(B(j)) ≠ 0, then restricting the predicate to the individual j does not change the conditional probability, as P(A | B) = P(A(j) | B(j)). However, there may be a predicate B' such that P(Vx(B'(x) → B(x))) = 1, but P(A || B') = r does not hold for any r (see figure 1).

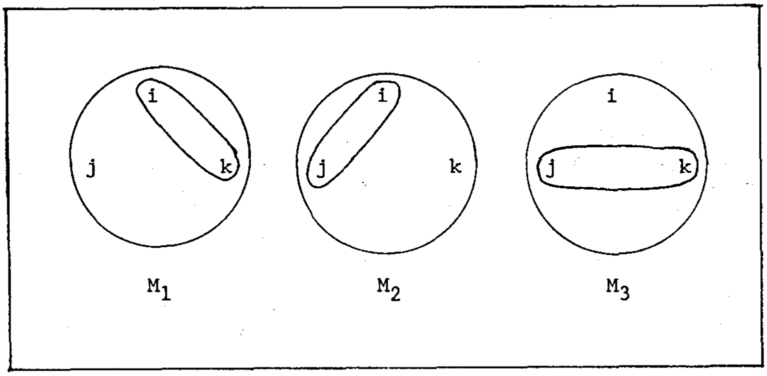

Figure 1: Let $P(M_1) = P(M_2) = P(M_3) = 1/3$, and let $B^{M}i = M_i$ $1 \leq i \leq 3$. Let $A^{M}i$ be represented by the figures inside M_i $1 \leq i \leq 3$. Then P(A || B) = 2/3. Let B' = B except that $B'^{M}3 = \emptyset$. P(A || B') = r does not hold for any r.

Consider now the inference:

$$2)\quad \begin{array}{l} P(A \,||\, B) \geq q \\ \underline{P(B(j)) \geq r} \\ P(A(j)) \geq rq \end{array}$$

The remark to theorem 1 implies that the conclusion of (2) is true whenever the premises are true. Hence, if rq is close to one we may consider (2) as a valid probabilistic explanation for the explanandum A(j). The premises do by themselves make the explanandum plausible. The following theorem shows that we cannot replace the major premise in (2) by the ordinary conditional probability $P(A \mid B)$.

Theorem 2: Let $0 < q, r < 1$ be rational numbers. Then there is a language L of finite domain with cardinality D with one-place predicates A and B and with probability model P such that $P(A\ B) = q$, $P(B(j)) = r$ and $P(A(j)) = 0$, for some individual j.

Proof: First choose positive integers a and b such that

3) $q = (1 - r)/((ra)/b + (1 - r))$,

and choose D such that $D \geq \{\max a,b\}+1$. Choose M, N ϵ M(L) satisfying:

$$|B^M| = a;\ |B^N| = b;\ j^M \epsilon\ B^M;\ j^N \not\epsilon\ B^N;\ A^M = \emptyset;\ A^N = B^N.$$

Define $P(M) = r$, $P(N) = 1 - r$. One easily verifies:

$$P(B(j)) = r$$
$$P(B) = ra/D + (1 - r)b/D$$
$$P(A \text{ and } B) = (1 - r)b/D$$

Using (3) it is easy to check that:

$$P(A \mid B) = q, \text{ and } P(A(j)) = 0. \ ///$$

3. Probabilistic Inference in MYCIN

The argument (2) bears a formal resemblance to the probabilistic inference schemes used in the so-called "production rule" expert systems. The best known system of this type is MYCIN (Buchanan and Shortliffe 1975) developed at the Stanford Heuristic Programming Project in collaboration with the Infectious Disease Group at the Stanford Medical School. MYCIN was designed to assist physicians in the diagnosis and treatment of certain kinds of bacterial infections.

A full discussion of MYCIN lies beyond the scope of this paper. A good summary can be found in O'Shea and Eisenstadt (1984), and a full discussion can be found in Buchanan and Shortliffe (1984).

The most distinctive feature of MYCIN's probabilistic inference schemes is its use of "certainty factors". Shortliffe and Buchanan define the certainty factor for hypothesis h given evidence e, $CF(h \mid e)$ as follows:

$$\begin{aligned} 4)\quad CF(h \mid e) &= 1 \text{ if } P(h) = 1 \\ &= -1 \text{ if } P(h) = 0 \\ &= (P(h \mid e)-P(h))/(1-P(h)) \text{ if } P(h \mid e) \geq P(h) \\ &= (P(h \mid e)-P(h))/P(h) \text{ if } P(h \mid e) < P(h). \end{aligned}$$

The following is an example of an inference rule, or production rule from MYCIN:

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

If: (1) the stain of the organism is gram positive, and
(2) the morphology of the organism is coccus, and
(3) the growth conformation of the organism is chains

Then: there is suggestive evidence (0.7) that the identity of the organism is streptococcus.

The numerical value for the certainty factor 0.7 is elicited from experts with the prompt: "On a scale from 1 to 10, how much certainty do you affix to this conclusion?" (Buchanan and Shortliffe 1975, p. 357). Buchanan and Shortliffe claim that the CF function captures the way experts reason from evidence. They claim that evidence which strengthens belief in hypothesis h does not necessarily weaken belief in not-h. They note, however, that at the beginning of a consultation P(h) is small for all h. In this case $CF(h | e)$ may be regarded as an "approximation" to $P(h | e)$, in their view. To be used in a rule, a certainty factor must be at least 0.2.

MYCIN also provides a mechanism for combining the certainty factors for a given conclusion h when this can be derived from different rules, with different evidential premises. Suppose we dispose over evidence e and e′, and we have rules with certainty factors $CF(h | e)$ and $CF(h | e')$. As we cannot in general calculate $P(h | e \text{ and } e')$ from $P(h | e)$ and $P(h | e\ )$ Buchanan and Shortliffe introduce the following "approximation technique (we restrict ourselves to the case where $CF(h | e) > 0$, and $CF(h | e') > 0$). The combined certainty factor $CCF(h | e \text{ and } e')$ is defined as:

5) $CCF(h | e \text{ and } e'){:}{-}CF(h | e) + CF(h | e') - CF(h | e)CF(h | e')$.

Adams (1976) showed that this rule and a similar rule for negative certainty factors are consistent with the probabilistic interpretation given previously (i.e., $CCF(h | e \text{ and } e') = CF(h | e \text{ and } e')$) if e and e′ are independent in h, in not-h and in the population as a whole.

Certainty factors are also attributed to singular statements, such as "the stain of the organism is gram positive". These factors may be estimated directly or derived from other rules. Rules of combination are also provided for deriving a certainty factor for the premise of a rule. Letting g and h denote clauses to which certainty factors have previously been assigned, then:

6) $CF(g \text{ and } h) = \min\{CF(g), CF(h)\}$
$CF(g \text{ or } h) = \max\{CF(g), CF(h)\}$.

Letting e denote the premise of a rule, the above combination rules enable us to calculate CF(e) from the certainty factors of the clauses in e. The certainty factor attaching to the conclusion of the rule when the premise is known with certainty factor CF(e) is taken to be the product of the certainty factor of the rule, $CF(h | e)$ and CF(e).

In 1977 the MYCIN inference mechanism was overhauled (see Buchanan and Shortliffe 1984, p. 216). The only significant change concerning the rules discussed here involved the combination rule (5), but as this will not play a large role in the discussion of the next section we shall not pursue these developments.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

MYCIN may be considered a very successful expert system, at least when "success" is defined in terms of passing a Turing test. Yu et al. (1979) present two evaluations by expert judges of MYCIN's treatment recommendations concerning respectively 15 cases of bacteraemia and 10 cases of meningitis. MYCIN's therapy recommendations were judged "unacceptable" in 27% and 35% of these cases respectively. The latter study also involved 8 recommendations by clinicians of varying seniority in a blind experiment. All eight human experts had a higher percentage of their recommendations judged "unacceptable" than MYCIN.

Finally, it should be noted that some of the people who have worked on MYCIN have distanced themselves from the probability interpretation of the certainty factors given above. Van Melle et al. (1981) claim that certainty factors are to be interpreted as "single numbers combining subjective probabilities and utilities. As such they represent the importance of the fact." (Cited in Spiegelhalter and Knill-Jones (1984, p. 41)). Gordon and Shortliffe (1984) have recently argued that the Dempster-Shafer theory of belief functions may yield a better representation of probabilistic reasoning than the theory presented here.

4. Reconstruction of MYCIN's Inference Engine

In this section we consider the problem of giving a formal representation of MYCIN's probabilistic reasoning. Such reasoning is almost always concerned with establishing the identity of an organism in a specific culture taken from a patient. It is natural then to let an individual name refer to an organism in a specific culture. It is also natural to consider a MYCIN rule as some sort of probabilistic conditional, and to consider a statement asserting the premise of a rule as a probabilistic singular statement. It is then apparent that the traditional formalism of probabilistic explanation as expressed in (1) is not appropriate, and the probabilistic semantics developed in Section (2) would seem to provide a natural formal setting for representing this type of reasoning.

The goal of this section is twofold. First we wish to determine whether a valid probabilistic inference scheme exists for the type of situation in which MYCIN operates. Second, we wish to determine whether the MYCIN scheme itself has a valid interpretation in terms of probability semantics. We content ourselves with an informal definition of validity for probabilistic inference schemes: *An inference scheme will be termed probabilistically valid if its conclusion holds in every probability model in which its premises hold.*

The remark to theorem 1 shows that the scheme:

2) $$\frac{\begin{array}{l} P(A \,||\, B) \geq q \\ P(B(j)) \geq r \end{array}}{P(A(j)) \geq qr}$$

is probabilistically valid. In order to determine whether this scheme is applicable, we must verify that the probabilistic conditional appearing as the major premise in (2) is appropriate, and we must show how the probability of B(j)' can be bounded from below.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

Given that the clauses in the premise of a MYCIN rule concern an individual organism, named j, we may interpret B(j) as the appropriate Boolean combination of these clauses. With the prompt:

> "What is the greatest lower bound for the probability that A(x), given that B(x)?"

we could solicit a value from experts which could be filled in for "q" in (2).

We must also show how P(B(j)) can be bounded from below. The min function in (6) won't work, as min {P(B(j)), P(B'(j))} is an *upper* bound for P(B(j) and B'(j)). However, the following will work:

7) $P(B(j) \text{ and } B'(j)) \geq P(B(j)) + P(B'(j)) - 1.$
$P(B(j) \text{ or } B'(j)) \geq \max\{P(B(j)), P(B'(j))\}$
$P(\text{not-}B(j)) = 1 - P(B(j))$

Whenever the clauses in B(j) can be inferred from other rules, (7) will allow a lower bound to be calculated for P(B(j)), which can then be used in (2).

These lower bounds are crude, which of course is appropriate when no further information on the dependence of the clauses is available. Any serious implementation of this mechanism in an expert system would involve "teaching" the system to extract global estimates of dependencies from its data base (or alternatively, to use its data to update default values of dependency). Such estimates would allow the bounds in (7) to be improved considerably. For example, if it were known that the clauses B(j) and B'(j) were not negatively correlated, then we could underestimate their conjunction via:

$$P(B(j) \text{ and } B'(j)) \geq P(B(j))P(B'(j)) \ .$$

A strategy for choosing rules would be. Look for rules with premises whose probability can be estimated efficiently from below.

The above scheme would seem to be perfectly suitable to the MYCIN type of situation. We must now consider what happens when a conclusion can be derived via two different inference rules. Suppose, in addition to the above rule, we also have:

$$\frac{\begin{array}{l} P(A \,\|\, C) \geq s \\ P(C(j)) \geq t \end{array}}{P(A(j)) \geq st}$$

As noted in Section 2, we cannot say anything in general about $P(A \,\|\, B \text{ and } C)$. The only "rule of combination" which can be applied for these two rules is:

$$P(A(j)) \geq \max\{rq, st\}.$$

We conclude that the above constitutes a valid inference procedure which is applicable to the MYCIN type situation.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

We now turn to the question whether MYCIN's own inference scheme can be reconstructed as a valid inference scheme in probability semantics. In other words, does the MYCIN inference scheme lead to probabilistically valid conclusions in every probability model? In order to carry out this analysis it is necessary to introduce some constraint on the relation between certainty factors and probabilities. The very name "certainty factor" suggests a monotonic relation to probabilities. Indeed, if hypothesis h is given a higher certainty factor than hypothesis h', users of MYCIN will undoubtably infer that h is more likely to be true than h'. However, the negative results which we reach below can be derived under much weaker assumption. It suffices to assume that there is some function f: [0,1] → [-1,1], taking probabilities into certainty factors and satisfying:

$f^{-1}(1) = 1$; $f^{-1}(-1) = 0$; and

f is the same for all probability models.

Under these two very weak assumptions we show that MYCIN's scheme cannot be probabilistically valid when the conditional probabilities in (4) are interpreted either as $P(A \mid B)$ or $P(A \parallel B)$. We consider each possibility in turn.

Using $P(A \mid B)$ in definition (4) and assuming $P(A \mid B) > P(A)$, we can represent a MYCIN inference as:

9)
$$\begin{array}{l} P(A \mid B) = q(1 - P(A)) + P(A) \\ P(B(j)) = r \\ \hline P(A(j)) \in f^{-1}(qf(r));\ 0.2 \leq q,\ f(r) \end{array}$$

Here, q is the certainty factor in the corresponding MYCIN rule. Theorem 2 shows that this scheme is invalid if q, f(r) < 1; we can find a probability model P satisfying the premises of (9) such that P(A(j)) = 0. This implies that qf(r) = -1. However, as q, f(r) 0.2, this is clearly impossible.

Next, we consider using $P(A \parallel B)$ in (4). We represent P(h) as P(A). Were we to use P(A(j)) instead of P(A), then the value of the certainty factor would depend on the individual J being considered, and this is clearly not the intention of the definition. This yields the following representation of a MYCIN inference:

10)
$$\begin{array}{l} P(A \parallel B) = q(1 - P(A)) + P(A) \\ P(B(j)) = r \\ \hline P(A(j)) \in f^{-1}(qf(r));\ 0.2 \leq q,\ f(r) \end{array}$$

Choose predicates A and B such that equality holds in the remark following Theorem 1. Then $P(A(j)) = P(B(j))P(A \parallel B)$. Writing a = P(A) and assuming q < 1, (10) implies:

11) $f(r(q(1 - a) + a)) = qf(r)$.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

Using the transformation $x = q(1-a) + a$, setting $r = 1$ and recalling that $f(1) = 1$, we find:

$$f(x) = q = (x - a)/(1 - a); \quad x < 1.$$

We see that the value of f at x must depend on a. To see that the MYCIN scheme cannot be probabilistically valid, it suffices to consider a second probability model P′ such that:

12) $(P'(A \,||\, B) - P'(A))/(1 - P'(A)) = q' \geq 0.2; \quad P'(B(j)) = 1,$

13) $a' = P'(A) \neq P(A)$

14) $q'(1 - a') + a' = q(1 - a) + a.$

MYCIN would now permit an inference of the form (10) with P′ replacing P. However, repeating the above argument with P′ replacing P would lead us to conclude:

$$(x - a')/(1 - a') = (x - a)/(1 - a),$$

which implies that $x = 1$. It follows that MYCIN's inference scheme is not probabilistically valid under any interpretation of the certainty factors in terms of probabilities which associates maximal certainty with probabilities one and zero and which is invariant for all probability models.

5. Conclusion

In concluding his critical analysis of the combination rule (5), Adams states:

> The empirical success of MYCIN using the model of Shortliffe and Buchanan stands in spite of theoretical objections of the types discussed in the preceding sections. It is probable that the model does not founder on the difficulties pointed out because in actual use the chains of reasoning are short and the hypotheses simple. However, there are many fields in which, because of its shortcomings, this model could not enjoy comparable success. (Adams 1976, p. 184-185)

The analysis in the preceding sections provides a further understanding of Adam's claim that MYCIN works well only if the chains of inference are short.

First of all, in spite of the fact that no valid reconstruction of the MYCIN inference scheme suggests itself, there is a valid inference scheme which is applicable to the MYCIN situations. This is the scheme embodied in (2), with combination rules (7) and (8). From combination rule (7) it is obvious that a premise involving several conjuncts will not survive unless the probability of each conjunct is high. This will obviously prevent long chains of inference where probability is lost at each step. The comparable MYCIN rule (6) may in principle permit longer chains of inference. However, if in fact the chains of inference are short, this effect may not be too strong.

Second, the prompt used by Buchanan and Shortliffe in soliciting certainty factors from experts may well produce the same, or roughly the same numerical results as the prompt appropriate to the valid inference scheme discussed in Section 4.

Finally, Buchanan and Shortliffe (1984, p. 219) note that MYCIN's diagnoses are extremely robust with respect to coarse graining the scale of the certainty factors. Distinguishing only 5 values for certainty factors did not lead to large differences in diagnosis or therapy recommendation.

In light of these facts, it would seem that the valid inference scheme proposed here could give a good explanation of MYCIN's success, and would allow a realistic appraisal of its potential in other domains.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press

References

Adams, J.B. (1976). "Probability Model of Medical Reasoning and the MYCIN Model." Mathematical Biosciences 32: 177-186.

Buchanan, B. and Shortliffe, E. (1975). "A Model of Inexact Reasoning in Medicine." Mathematical Biosciences 23: 351-379.

- - - - - - - - - - - - - - - - - - . (1984). Rule-Based Expert Systems. Reading, Mass: Addison-Wesley.

Cooke, R.M. (1981). "A Paradox in Hempel's Criterion of Maximal Specificity." Philosophy of Science 48: 327-328.

Gordon J. and Shortliffe, E. (1984). "The Dempster-Shafer Theory of Evidence." In Buchanan and Shortliffe (1984).

Hempel, C. (1968). "Maximal Specificity and Lawlikeness in Probabilistic Explanation." Philosophy of Science 35: 116-133.

Los, J. (1963). "Semantic Representation of the Probability of Formulas in Formalized Theories." Studia Loika 14: 183-196.

O'Shea, T. and Eisenstadt, M. (1984). Artificial Intelligence: Tools, Techniques and Applications. New York: Harper and Row.

Spiegelhalter, D. and Knill-Jones, R. (1984). "Statistical and Knowledge-based Approaches to Clinical Decision-support Systems, with an Application in Gastroenterology." Journal of the Royal Statistical Society, Series A 147: 35-77.

Van Melle, W.; Scott, A.; Bennet, J.; and Peairs, M. (1981). The Manual Rep. HPP-81-16. Stanford University, Computer Science Department.

Yu, V.; Fagan, L.; Wraith, S.; Clancey, W.; Scott A.; Hannigan, J.; Blum, R.; Buchanan, B.; and Cohen, S. (1979). "Antimicrobial Selection by a Computer: a blinded evaluation by infectious disease experts." Journal American Medical Association 242: 1279-1282.

https://doi.org/10.1086/psaprocbienmeetp.1986.1.193141 Published online by Cambridge University Press