

12

Probabilistic Reasoning and Certainty Factors

J. Barclay Adams

The development of automated assistance for medical diagnosis and decision making is an area of both theoretical and practical interest. Of methods for utilizing evidence to select diagnoses or decisions, probability theory has the firmest appeal. Probability theory in the form of Bayes' Theorem has been used by a number of workers (Ross, 1972). Notable among recent developments are those of de Dombal and coworkers (de Dombal, 1973; de Dombal et al., 1974; 1975) and Pipberger and coworkers (Pipberger et al., 1975). The usefulness of Bayes' Theorem is limited by practical difficulties, principally the lack of data adequate to estimate accurately the *a priori* and conditional probabilities used in the theorem. One attempt to mitigate this problem has been to assume statistical independence among various pieces of evidence. How seriously this approximation affects results is often unclear, and correction mechanisms have been explored (Ross, 1972; Norusis and Jacquez, 1975a; 1975b). Even the independence assumption requires an unmanageable number of estimates of probabilities for most applications with realistic complexity. To circumvent this problem, some have tried to elicit estimates of probabilities directly from experienced physicians (Gorry, 1973; Ginsberg, 1971; Gustafson et al., 1971), while others have turned from the use of Bayes' Theorem and probability theory to the use of discriminant analysis (Ross, 1972) and nonprobabilistic methods (Scheinok and Rinaldo, 1971; Cumberbatch and Heaps, 1973; Cumberbatch et al., 1974; Glesser and Collen, 1972).

Shortliffe and Buchanan (1975) have offered a model of inexact reasoning in medicine used in the MYCIN system (Chapter 11). Their model

This chapter is a shortened and edited version of a paper appearing in *Mathematical Biosciences* 32: 177-186 (1976). Copyright © 1976 by *Mathematical Biosciences*. All rights reserved. Used with permission.

uses estimates provided by expert physicians that reflect the tendency of a piece of evidence to prove or disprove a given hypothesis. Because of the highly promising nature of the MYCIN system, this model deserves examination. Shortliffe and Buchanan conceived their system purely on intuitive grounds and assert that it is an alternative to probability theory. I shall show below that a substantial part of this model can be derived from and is equivalent to probability theory with the assumption of statistical independence. In Section 12.1 I first review a simple probability model and discuss some of its limitations.

12.1 A Simple Probability Model

Consider a finite population of n members. Members of the population may possess one or more of several properties that define subpopulations or sets. Properties of interest might be e_1 or e_2 , which might be evidence for or against a disease, and h , a certain disease state or other hypothesis about an individual. The number of individuals with a certain property, say e , will be denoted $n(e)$, and the number with both of two properties e_1 and e_2 will be denoted $n(e_1 \& e_2)$. Probabilities are taken as ratios of numbers of individuals. From the observation that:

$$\frac{n(e \& h)}{n(e)} \cdot \frac{n}{n(h)} = \frac{n(e \& h)}{n(h)} \cdot \frac{n}{n(e)}$$

a convenient form of Bayes' Theorem follows immediately:

$$\frac{P(h|e)}{P(h)} = \frac{P(e|h)}{P(e)}$$

Now consider the case in which two pieces of evidence e_1 and e_2 bear on a hypothesis or disease state h . Let us make the assumptions that these pieces of evidence are independent both in the population as a whole and in the subpopulation with h ; that is:

$$\frac{n(e_1 \& e_2)}{n} = \frac{n(e_1)}{n} \cdot \frac{n(e_2)}{n} \quad (1)$$

and

$$\frac{n(e_1 \& e_2 \& h)}{n(h)} = \frac{n(e_1 \& h)}{n(h)} \cdot \frac{n(e_2 \& h)}{n(h)} \quad (2)$$

or

$$P(e_1 \ \& \ e_2) = P(e_1)P(e_2) \tag{3}$$

and

$$P(e_1 \ \& \ e_2|h) = P(e_1|h)P(e_2|h) \tag{4}$$

With these the right-hand side of Bayes' Theorem becomes

$$\frac{P(e_1 \ \& \ e_2|h)}{P(e_1 \ \& \ e_2)} = \frac{P(e_1|h)}{P(e_1)} \cdot \frac{P(e_2|h)}{P(e_2)} \tag{5}$$

and, because of this factoring, the right-hand side is computationally simple.

Now, because of the dearth of empirical data to estimate probabilities, suppose we were to ask experts to estimate the probabilities subjectively. We could ask for estimates of the ratios $P(e_i|h)/P(e_i)$ and $P(h)$, and from these compute $P(h|e_i \ \& \ e_2 \ \& \ \dots \ \& \ e_n)$. The ratios $P(e_i|h)/P(e_i)$ must be in the range $[0,1/P(h)]$. Most physicians are not accustomed to thinking of diseases and evidence in terms of probability ratios. They would more willingly attempt to quantitate their intuition by first deciding whether a piece of evidence tends to prove or disprove a hypothesis and then assigning a parameter on a scale of 0 to 10 as a measure of the weight or strength of the evidence. One way to translate this parameterization into an "estimate" of a probability ratio is the following. Divide the intuitive parameter by 10, yielding a new parameter, which for evidence favoring the hypothesis will be called MB, the physician's measure of belief, and for evidence against the hypothesis will be called MD, the physician's measure of disbelief. Both MB and MD are in the range $[0,1]$ and have the value 0 when the evidence has no bearing on the hypothesis. The value 1 for MB $[h,e]$ means that all individuals with e have h . The value 1 for MD $[h,e]$ means that no individual with e has h . From these physician-estimated parameters we derive the corresponding probability ratios in the following way. For evidence against the hypothesis we simply take

$$\frac{P(e|h)}{P(e)} = 1 - \text{MD}[h,e] \tag{6}$$

For evidence favoring the hypothesis we use a similar construct by taking the evidence as against the negation of the hypothesis, i.e., by considering the subpopulation of individuals who do not have h , denoted $\neg h$. So we construct the ratio of probabilities using MB:

$$\frac{P(e|\neg h)}{P(e)} = 1 - \text{MB}[h,e] \tag{7}$$

Now, to continue the parallel, we write Bayes' Theorem for two pieces of evidence favoring a hypothesis:

$$\frac{P(\neg h|e_1 \& e_2)}{P(\neg h)} = \frac{P(e_1 \& e_2|\neg h)P(e_1 \& e_2)}{P(e_1 \& e_2)} \quad (8)$$

with

$$\frac{P(e_1 \& e_2|h)}{P(e_1 \& e_2)} = \frac{P(e_1|\neg h)}{P(e_1)} \cdot \frac{P(e_2|\neg h)}{P(e_2)} \quad (9)$$

where, for the last equality, independence of e_1 and e_2 in $\neg h$ is assumed. By using the identities

$$P(h) + P(\neg h) = 1 \quad (10)$$

$$P(h|e) + P(\neg h|e) = 1 \quad (11)$$

one then has a computationally simple way of serially adjusting the probability of a hypothesis with new evidence against the hypothesis:

$$P(h|e'') = \frac{P(e_i|h)}{P(e_i)} \cdot P(h|e') \quad (12)$$

or new evidence favoring the hypothesis:

$$P(h|e'') = 1 - \frac{P(e_i|\neg h)}{P(e_i)} \cdot [1 - P(h|e')] \quad (13)$$

where e_i is the new evidence, e'' is the total evidence after the introduction of e_i , and e' is the evidence before the new evidence is introduced [note that $P(h|e')=P(h)$ before any evidence is introduced]. Alternatively, one could combine all elements of evidence against a hypothesis simply by using independence as in Equation (5) and separately combine all elements of evidence favoring a hypothesis by using Equation (9), and then use Equations (12) and (13) once.

The attractive computational simplicity of this scheme is vitiated by the restrictive nature of the independence assumptions made in deriving it. The MB's and MD's for different pieces of evidence cannot be chosen arbitrarily and independently. This can be clearly seen in the following simple theorem. If e_1 and e_2 are independent both in the whole population and in the subpopulation with property h , then

$$P(h|e_1)P(h|e_2) = P(h|e_1 \& e_2)P(h) \quad (14)$$

This follows from dividing Equation (2) by Equation (1). The nature of restrictions placed on the probabilities can be seen from the limiting case in which all members of e_1 are in h . In that case, $P(h|e_1) = P(h|e_1 \ \& \ e_2) = 1$, so $P(h|e_2) = P(h)$; that is, if some piece of evidence is absolutely diagnostic of an illness, then any evidence that is independent can have no diagnostic value. This special case of the theorem was noted in a paper of Warner et al. (1961). Restrictions this forces on the MB's can be further demonstrated by the following example. We write Bayes' Theorem with the independence assumption as follows:

$$\frac{P(e_1|h)}{P(e_1)} \cdot \frac{P(e_2|h)}{P(e_2)} = \frac{P(h|e_1 \ \& \ e_2)}{P(h)} \tag{15}$$

Consider the case of two pieces of evidence that favor the hypothesis. Using Equations (6), (10), and (11), one can express $P(e|h)/P(e)$ in terms of MB as follows:

$$\frac{P(e|h)}{P(e)} = 1 + \left(\frac{1}{P(h)} - 1 \right) \text{MB}[h,e] \tag{16}$$

Using this form and the fact that $P(h|e_1 \ \& \ e_2) \leq 1$, we get from Equation (15)

$$\left\{ 1 + \left(\frac{1}{P(h)} - 1 \right) \text{MB}[h,e_1] \right\} \left\{ 1 + \left(\frac{1}{P(h)} - 1 \right) \text{MB}[h,e_2] \right\} \leq \frac{1}{P(h)} \tag{17}$$

This is not satisfied for all values of the MB's; e.g., if $P(h) = 1/11$ and $\text{MB}[h,e_1] = 0.7$, then we must choose the narrow range $\text{MB}[h,e_2] \leq 0.035$ to satisfy the inequality. Most workers in this field assume that elements of evidence are statistically independent only within each of a complete set of mutually exclusive subpopulations and not in the population as a whole; thus the properties of (14) and (15) do not hold. Occasionally, writers have implicitly made the stronger assumption of independence in the whole space (Slovic et al., 1971).

12.2 The MYCIN Model

The model developed by Shortliffe and Buchanan is in part equivalent to that in Section 12.1. They introduce quantities $\text{MB}[h,e]$ and $\text{MD}[h,e]$, which are identical to those we have defined above (and were the reason for selecting our choice of parameterization). They postulate rules for com-

binning $MB[h, e_1]$ with $MB[h, e_2]$ to yield $MB[h, e_1 \& e_2]$ and similar rules for MD. With one exception discussed below, these rules need not be postulated because they are equivalent to, and can be derived from, the method of combining probability ratios under the assumption of independence used in the previous section. For example, the rule for MD's is derived as follows by using Equation (5):

$$1 - MD[h, e_1 \& e_2] = \frac{P(e_1 \& e_2|h)}{P(e_1 \& e_2)} = \frac{P(e_1|h)}{P(e_1)} \cdot \frac{P(e_2|h)}{P(e_2)} \quad (18)$$

or

$$1 - MD[h, e_1 \& e_2] = (1 - MD[h, e_1])(1 - MD[h, e_2]) \quad (19)$$

which is an algebraic rearrangement of the rule postulated in their paper. A similar construct holds for MB. The exceptional case in the MYCIN model is one in which a piece of evidence proves a hypothesis (all with e_1 have h). As noted in the previous section, this case excludes the possibility of other independent diagnostically meaningful evidence. In the MYCIN model, if e proves h , then one sets MD equal to zero for the combined evidence. A similar assumption is introduced for the case that evidence disproves a hypothesis. To maintain internal consistency the MB's and MD's must be subject to the restrictions discussed in Section 12.1. This important fact is not noted in the work of Shortliffe and Buchanan.

Two other properties are assumed for the MB's and MD's by Shortliffe and Buchanan. The extent or importance of the use of these assumptions in the employment of their model is not clear, but does not seem great. One concerns the conjunction of hypotheses h_1 and h_2 , for which they assume

$$MB[h_1 \& h_2, e] = \min(MB[h_1, e], MB[h_2, e]) \quad (20)$$

$$MD[h_1 \& h_2, e] = \max(MD[h_1, e], MD[h_2, e]) \quad (21)$$

Unstated are strong restrictive assumptions about the relationship of h_1 and h_2 . As an extreme example, suppose that h_1 and h_2 are mutually exclusive; then the conjunction $h_1 \& h_2$ is false (has probability zero) no matter what the evidence, and the assumptions on the conjunction of hypotheses would be unreasonable. In the context of the probability model of Section 12.1, one can derive a relationship

$$\frac{P(h_1 \& h_2|e)}{P(h_1 \& h_2)} = \frac{P(h_1|e)}{P(h_1)} \cdot \frac{P(h_2|e)}{P(h_2)} \quad (22)$$

only by making strong assumptions on the independence of h_1 and h_2 .

A pair of further assumptions made by Shortliffe and Buchanan concerns the disjunction of two hypotheses, denoted $h_1 \vee h_2$. These are

$$MB[h_1 \vee h_2, e] = \max(MB[h_1, e], MB[h_2, e]) \tag{23}$$

$$MD[h_1 \vee h_2, e] = \min(MD[h_1, e], MD[h_2, e]) \tag{24}$$

Again these contain unstated assumptions about the relationship of h_1 and h_2 . If, for example, h_1 and h_2 are mutually exclusive and each has a probability of being true, then the disjunction $h_1 \vee h_2$ should be more likely or probable or confirmed than either h_1 or h_2 . Expressions for $P(e|h_1 \vee h_2)/P(e)$ can be derived in probability theory, but they have no compact or perspicuous form.

The MYCIN model combines separately all evidence favoring a hypothesis to give $MB[h, e_f]$, where $e_f = e_{f1} \& e_{f2} \& \dots \& e_{fn}$, the intersection of all elements of evidence favoring hypothesis h . Similarly, all elements against a hypothesis are combined to give $MD[h, e_a]$. By Bayes' Theorem these provide measures of $P(h|e_f)/P(h)$ and $P(h|e_a)/P(h)$. These could be combined using the probability theory outlined in Section 12.1 to give $P(h|e_f \& e_a)/P(h)$, an estimate of the change of the probability due to the evidence. However, it is at this point that the MYCIN model departs from standard probability theory. Shortliffe and Buchanan combine the MB with the MD by defining a certainty factor to be

$$CF[h, e_f \& e_a] = MB[h, e_f] - MD[h, e_a] \tag{25}$$

The certainty factor is used in two ways. One is to rank hypotheses to select those for further action. The other is as a weighting factor for the credibility of a hypothesis h , which is supposed by an intermediate hypothesis i , which in turn is supported by evidence e . The appropriateness of CF for each of these roles will be examined.

One of the uses of CF is to rank hypotheses. Because $CF[h, e]$ does not correspond to the probability of h given e , it is not difficult to give examples in which, of two hypotheses, the one with the lower probability would have the higher certainty factor, or CF. For example, consider two hypotheses h_1 and h_2 and some body of evidence e that tends to confirm both hypotheses. Suppose that the *a priori* probabilities were such that $P(h_1) \gg P(h_2)$ and $P(h_1|e) > P(h_2|e)$; it is possible that $CF[h_1, e] < CF[h_2, e]$. For example, if $P(h_1) = 0.8$, $P(h_2) = 0.2$, $P(h_1|e) = 0.9$, $P(h_2|e) = 0.8$, then $CF[h_1, e] = 0.5$ and $CF[h_2, e] = 0.75$. This failure to rank according to probabilities is an undesirable feature of CF. It would be possible to avoid it if it were assumed that all *a priori* probabilities were equal.

The weighting role for CF is suggested by the intuitive notion that in a chain of reasoning, if e implies i with probability $P(i|e)$, and i , if true, implies h with probability $P(h|i)$, then

$$P(h|e) = P(h|i)P(i|e) \tag{26}$$

This is not true in general; however, a set of assumptions can be identified under which it will be true. Suppose the population with property h is contained in the set with i , and the set with i is contained in the set with e . This may be expressed as

$$n(h \ \& \ i) = n(h) \quad n(i \ \& \ e) = n(i) \quad n(h \ \& \ e) = n(h) \quad (27)$$

These allow us to write

$$\frac{n(h \ \& \ e)}{n(e)} = \frac{n(h \ \& \ i)}{n(i)} \cdot \frac{n(i \ \& \ e)}{n} \quad (28)$$

which is the desired result in numerical form. The proposal of Shortliffe and Buchanan, which may be written as

$$\text{MB}[h,e] = \text{MB}[h,i] \max(0, \text{CF}[i,e]) \quad (29)$$

$$\text{MD}[h,e] = \text{MD}[h,i] \max(0, \text{CF}[i,e]) \quad (30)$$

is not true in general under the assumptions of (27) or any other natural set, as may be demonstrated by substitution into these relationships of the definitions of MB, MD, and CF.

12.3 Conclusions

The simple model of Section 12.1 is attractive because it is computationally simple and apparently lends itself to convenient estimation of parameters by experts. The weakness of the system is the inobvious interdependence restriction placed on the estimation of parameters by the assumptions of independence. The MYCIN model is equivalent in part to the simple probability model presented and suffers from the same subtle restrictions on parameter estimation if it is to remain internally consistent.

The ultimate measure of success in models of medical reasoning of this sort, which attempt to mimic physicians, is the closeness of their approach to perfect imitation of experts in the field. 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.

The fact that in trying to create an alternative to probability theory or reasoning Shortliffe and Buchanan duplicated the use of standard theory

demonstrates the difficulty of creating a useful and internally consistent system that is not isomorphic to a portion of probability theory. In proposing such a system, a careful delineation of its relationship to conventional probability theory can contribute to an understanding and clear exposition of its assumptions and approximations. It thereby allows tests of whether these are satisfied in the proposed field of use.