Probabilistic Interpretations for MYCIN’s Certainty Factors

David Heckerman

Medical Computer Science Group, TC135, Stanford University Medical Center

Abstract

We propose a redefinition of the quantities called certainty factors used by MYCIN to manage uncertainty. This definition is based on the desiderata of certainty factors given in the original work. It is emphasized that this definition reflects the notion that certainty factors represent changes in belief about hypotheses given evidence as opposed to absolute beliefs. It is shown that this redefinition accommodates an infinite number of probabilistic interpretations. In constructing the interpretations, insight is provided into the assumptions made when propagating certainty factors through an inference net. For example, it is shown that all evidence which bears directly on a hypothesis must be conditionally independent on the hypothesis and its negation. Also, this analysis demonstrates that certainty factors must be elicited from experts in a precise fashion and a straightforward method for doing so is presented. Improvements in the methods currently used to propagate certainty factors through an inference net will also be suggested.

Introduction

Many AI researchers have created ad hoc strategies for reasoning under uncertainty. They have invariably explained their approaches as necessary because of limitations of probability theory. Many familiar with probability theory, however, argue that ad hoc methods for dealing with uncertainty are unnecessary and that probabilistic methods are sufficient in all cases. In this paper, we will focus on the MYCIN certainty factor model, a purportedly ad hoc method for managing uncertainty that has seen wide use in rule-based expert systems. Shortliffe and Buchanan, creators of the model, justify the use of certainty factors by demonstrating system performance equal to that of experts in the field. However, since they failed to provide a precise characterization of certainty factors, a good deal of mystery and confusion about the numbers has been generated. In this paper, we will clear up much of this mystery by showing that there are probabilistic interpretations for certainty factors. By demonstrating an equivalence between certainty factors and probabilities we will also lend support to the arguments that probability theory is sufficient for reasoning under uncertainty.

Previous Work

This work is not the first to attempt to gain a better understanding of certainty factors. Adams demonstrated several problems with the original definition of certainty factors. In an unpublished internal memo, Duda, Hart, Nilsson and Shortliffe proposed several new putative probabilistic interpretations but admittedly failed to acquire significant new insights into the certainty factor model. In fact, none of their interpretations fulfill the criteria for a probabilistic interpretation proposed below. Later, Duda demonstrated a strong similarity between the certainty factor model and the pseudo-probabilistic method used in the expert system PROSPECTOR to manage uncertainty. We will consider this similarity in a new light. Finally, Hjekk examined the algebraic properties of certainty factors without attempting a probabilistic interpretation. The significance of his work within the current context will be discussed.

MYCIN’s Certainty Factors

Before examining probabilistic interpretations for MYCIN’s certainty factors, we present a simplified overview of the quantities. MYCIN’s knowledge is stored as rules. Rules are of the form IF evidence THEN hypothesis. In this paper, rules will be represented by

\[ E \rightarrow \rightarrow \rightarrow H \]

where H is a hypothesis and E is evidence relating to the hypothesis. For example, a hypothesis might be that the organism Staphylococcus is infecting a patient while evidence for this hypothesis might be that the organism is gram-positive and grows in clumps.

In medicine, relationships between evidence and hypotheses are often uncertain. In order to accommodate these non-deterministic relationships, MYCIN uses certainty factors. To each rule, a certainty factor is attached which represents the change in belief about a hypothesis given some evidence. In this paper, we write

\[ CF(H, E) \]

\[ E \rightarrow \rightarrow \rightarrow H \]

Certainty factors range between -1 and 1. Positive numbers correspond to an increase in belief in a hypothesis while negative quantities correspond to a decrease in belief. Note that certainty factors do not correspond to measures of absolute belief. This has been a source of confusion with respect to certainty factors as well as other measures of uncertainty.

In MYCIN, it is possible that several pieces of evidence bear on the same hypothesis. It is also possible for a hypothesis to serve as evidence for another hypothesis. An interconnected set of hypotheses and evidence is called an inference network. Figure 1 depicts one possible net. In the figure, \( S_i \) is a statement which serves as either evidence or a hypothesis or sometimes both.
Given that belief that he is an extrovert is $.8$, if he likes to make solo backpacking trips, the change in belief that he is an extrovert is -.5. Furthermore, if a person is an extrovert, then the change in belief that he participates in some type of social work is $.4$. To compute the change in belief that a person is an extrovert given that he likes to go to parties and also likes to make solo backpacking trips, we use the parallel combination function (1) with $CF(H,E_1) = .8$ and $CF(H,E_2) = .4$ giving $CF(H,E_1E_2) = .8 \times .4 / (1 - .5) = .6$. To compute the change in belief that a person does some type of social work given that he likes to go to parties and also likes to make solo backpacking trips, we use the sequential combination function (2) with $CF(E,E') = .6$ (the result of parallel combination) and $CF(H,E) = .4$ giving $CF(H,E') = (.6 \times .4) = .24$.

This decomposition approach to propagation is straightforward when an inference net has a tree structure. However, MYCIN’s inference net is not a tree and additional ad hoc procedures were introduced to handle complexities such as multiple paths from evidence to hypothesis. To simplify this discussion, we will only consider tree networks.

**Overview of Interpretation development**

As mentioned above, the certainty factor $CF(H,E)$ represents the change in belief of a hypothesis $H$ based on the evidence $E$. In the original work, the creators of certainty factors presented this basic notion and then immediately proposed a formal definition of certainty factors in terms of probabilities. In somewhat simplified form and in our notation, Shortliffe and Buchanan gave as the definition of $CF(H,E)$:

$$
CF(H,E) = \begin{cases} 
\frac{p(H|E) - p(H)}{1 - p(H)} & p(H|E) > p(H) \\
1 & p(H|E) = p(H) \\
\frac{p(H) - p(H|E)}{1 - p(H)} & p(H|E) < p(H) 
\end{cases}
$$

Given this definition, however, they could not derive exact formulae for combining certainty factors (we will give one reason for this difficulty in a later section). Instead, they proposed "approximate" combination functions. They justified these functions by showing that they satisfied certain properties or "desiderata" which are consistent with the basic notion of certainty factors. For example, one of the desiderata is that parallel combination must not depend on the order in which the evidence is combined. They extended this definition of certainty factors while the term "CF" will refer to certainty factors that are combined with the MYCIN functions. The symbol $\Delta$ was chosen to emphasize that the quantity represents a belief update. A second distinction is made between certainty factors in general and particular probabilistic ones.

---

*Pearl* gives a detailed discussion of belief propagation through more complicated networks.
interpretations that may be ascribed to them. To highlight this distinction, a subscripted \( \Delta \), e.g., \( \Delta_1 \), will refer to a particular probabilistic interpretation.

**Formal properties of certainty factors**

We begin the construction of probabilistic interpretations for certainty factors by closely examining the desiderata or defining axioms. To understand these axioms, it is important to remember that \( \Delta(H,E) \) represents the change in belief of hypothesis \( H \) given evidence \( E \). In other words, \( \Delta \) is a belief update. The defining axioms fall into three categories: (1) basic properties not related to the combination functions, (2) properties of parallel combination, and (3) properties of sequential combination. Since sequential combination is somewhat more complicated than parallel combination, we will first discuss categories one and two and return to category three later.

**Basic axioms**

There are only two basic axioms of certainty factors. The first states that \( \Delta's \) are real numbers in the interval \([-1,1] \). The second states that \( \Delta's \) are ordered by the usual "<" relation on real numbers; that is, it makes sense to say that one certainty factor is less than another.

**Axioms of parallel combination**

The first axiom of parallel combination was not made explicit in the original work. Consider two rules with the same hypothesis \( H \):

\[
\begin{align*}
E \quad & \quad \rightarrow \quad H \\
E' \quad & \quad \rightarrow \quad H
\end{align*}
\]

It is possible that the update \( \Delta(H,E) \) may depend on whether or not \( E' \) is known. In general, the update \( \Delta(H,E) \) may depend on evidence for \( H \) already known. Therefore, certainty factors logically must have three arguments: \( H, E \) and \( e \) where \( e \) denotes prior evidence for \( H \). However, the creators of certainty factors omitted the third argument, implicitly assuming that an update is independent of previous updates. This was done so that rules would retain a modular character. Formally, the first axiom of parallel combination is

\[
\Delta(h, e) = \Delta(h, e, \emptyset) = \Delta(h, E)
\]

for every \( e \) which does not entail \( E \). This axiom will be called the modularity axiom. We will see later that it corresponds to the assumption of conditional independence.

As mentioned above, the parallel combination of two pieces of evidence \( E_1 \) and \( E_2 \) with respect to a hypothesis \( H \) is denoted \( \Delta(H, E_1, E_2) \). Given this notation, \( \Delta's \) are required to be associative

\[
\Delta(h, (E_1, E_2)) = \Delta(h, E_1, (E_2))
\]

and commutative

\[
\Delta(h, (E_1, E_2)) = \Delta(h, (E_2, E_1))
\]

In other words, updating belief should not depend on how evidence is grouped or on the order in which it combined. It is required that parallel combination preserves the ordering of updates. Suppose we have one update greater than another; that is \( \Delta(H, E_1, E_2) \geq \Delta(H, E_2, E_3) \). When we combine each of these with a third update \( \Delta(H, E_3) \), we want the two resulting updates to have the same relationship. Namely, if \( \Delta(H, E_1) > \Delta(H, E_2) \) then \( \Delta(H, E_1, E_2) > \Delta(H, E_2, E_3) \)

\[
\text{if } \Delta(H, E_1) > \Delta(H, E_2) \text{ then } \Delta(H, E_1, E_2) > \Delta(H, E_2, E_3)
\]

Another axiom of parallel combination states that 0 is a special certainty factor. In particular, if \( \Delta(H, E) = 0 \), the evidence \( E \) does not change the belief in hypothesis \( H \). This corresponds to a "no information" situation. In our notation,

\[
\text{if } \Delta(H, E_1) = 0 \text{ then } \Delta(H, E_1, E_2) = \Delta(H, E_2)
\]

for all \( E_2 \). In other words, 0 is the identity element with respect to parallel combination.

The next property of certainty factors we will discuss was not considered a desideratum in the original work. Rather, it was shown to follow from the definition of the MYCIN parallel combination. We will upgrade this property to an axiom. Informally the axiom states that if \( \Delta(H,E) \) corresponds to the increase in belief of \( H \) given \( E \), then \( -\Delta(H,E) \) should correspond to an equal and opposite decrease in belief. Formally,

\[
\text{if } \Delta(H, E_1) = -\Delta(H, E_2) \text{ then } \Delta(H, E_1, E_2) = 0
\]

That is, every certainty factor has an inverse with respect to parallel combination and it's inverse is just the certainty factor with opposite sign.

Finally, combination of "extreme" certainty factors must be consistent with the notion that a \( \Delta(H,E) \) of 1 corresponds to the case where \( E \) proves \( H \) and a \( \Delta(H,E) \) of -1 corresponds to the situation where \( H \) is disproved by \( E \).

\[
\text{if } \Delta(H, E_1) = 1 \text{ and } \Delta(H, E_2) < -1 \text{ then } \Delta(H, E_1, E_2) = 1
\]

\[
\text{if } \Delta(H, E_1) = -1 \text{ and } \Delta(H, E_2) > 1 \text{ then } \Delta(H, E_1, E_2) = -1
\]

\[
\text{if } \Delta(H, E_1) = 1 \text{ and } \Delta(H, E_2) = -1 \text{ then } \Delta(H, E_1, E_2)
\]

is undefined.

The last case corresponds to conflicting evidence where \( E_1 \) proves \( H \) while \( E_2 \) disproves \( H \).

**Requirements of a probabilistic interpretation**

Intuitively, a probabilistic interpretation for certainty factors is a combination of probabilistic quantities, in the spirit of (3), which satisfies the defining axioms. In this section, we will formalize this notion.

First, we require that there is some function \( f \) such that

\[
\Delta(h, E) = f(p(H|E), p(H))
\]

where \( p(H|E) \) denotes the conditional probability of \( H \) given \( E \), the posterior probability of \( H \), and \( p(H) \) denotes the probability of \( H \) before \( E \) is known, the prior probability. The only restriction we place on \( f \) here is that it be a smooth function.

Later we will see that the axioms of certainty factors place much tighter restrictions on \( f \). For later discussion, it is useful to state the above requirement with explicit reference to prior evidence:

\[
\Delta(h, E, e) = f(p(H|E), p(H|e))
\]

In addition to the above basic requirement, a probabilistic interpretation should be order preserving. In other words, suppose one update is greater than another. Given the same prior probability, the posterior probability corresponding to the first update should be greater than the posterior corresponding to the second. More formally,

\[
\Delta(h, E_1) < \Delta(h, E_2) \text{ if and only if } p(H|E_1) < p(H|E_2)
\]

with \( p(H) \) fixed.

Finally, a probabilistic interpretation should be consistent with the notion that \( \Delta(H,E) = 1 \) when \( H \) follows from \( E \) and \( \Delta(H,E) = -1 \) when \( E \) disproves \( H \).

\[
\Delta(h, E) = 1 \text{ if } p(H|E) = 1 \text{ and } p(H) < 0
\]

\[
\Delta(h, E) = 0 \text{ if } p(H|E) = 0 \text{ and } p(H) < 0
\]

*Precisely, the first derivative of \( f \) must exist and be continuous.*
The original interpretation of certainty factors

In this section, we examine the original definition (3) of certainty factors. Here, we view (3) as a probabilistic interpretation for certainty factors rather than a definition. To emphasize this we write

\[ \Delta_{\text{orig}}(H,E) = \begin{cases} 0 & \text{if } p(H) = p(H|E) \\ p(H|E) - p(H) & \text{if } p(H) > p(H|E) \\ 1 - p(H|E) & \text{if } p(H) < p(H|E) \end{cases} \quad (16) \]

There are several problems with this putative interpretation. Here, it is shown that (16) violates the axioms of certainty factors in this space range from -1 to 1. Hence, the original interpretation leads to the unappealing result that the belief in a hypothesis is dependent on the order in which evidence for the hypothesis is considered.

A probabilistic interpretation

The source of non-commutativity can be traced to the lack of symmetry in \( \Delta_{\text{orig}} \). From (15) we see that positive evidence is measured relative to the distance between 1 and \( p(H) \), while negative evidence is measured relative to the distance between 0 and \( p(H) \). In order to avoid this asymmetry, it seems reasonable to map \( p(H|E) \) and \( p(H) \) into the interval \((-1, 1)\). This gives some "elbow room" in which to combine \( \Delta \)'s without asymmetries. In fact, the obvious choice for a "certainty factor" in the infinite space is simply the distance between the mapped points of the posterior and prior probabilities. In this case, we can combine these certainty factors by adding them. However, there is now the problem that certainty factors in this space range from -\( \infty \) to \( \infty \). To fix this problem, we map these numbers into the interval \((-1, 1)\) with a function that preserves the \( \Delta \) properties. This is shown diagrammatically in Figure 2. Note that we are explicitly representing prior evidence \( e \) for reasons that will become clear shortly. However, this is not necessary given the modularity axiom of \( \Delta \)'. (4).

Referring to figure 2, the function \( F \) maps the posterior and prior probabilities into the interval \((-\infty, \infty)\). Certainty factors in the this infinite space, denoted \( \Delta^{\infty} \), are then constructed by subtracting the mapped posterior from the mapped prior

\[ \Delta^{\infty}(H,E,e) = F(p(H|E)) - F(p(H|e)) \quad (17) \]

Finally, the function \( G \) maps the \( \Delta^{\infty} \)'s into the interval \((-1, 1)\). That is,

\[ G(H,E,e) = G(\Delta^{\infty}(H,E,e)) \]

We cannot choose the functions \( F \) and \( G \) arbitrarily because the axioms of certainty factors and the requirements for a probabilistic interpretation must be respected. Later we will see that there is no freedom in the choice for \( F \). In particular,

\[ F(x) = \ln(x/(2-x)) \quad (18) \]

There is more freedom in the choice for \( G \) however. We will see that any \( G \) which satisfies

\[ G(e) = G(-e) \]

\[ G(\infty) = 1 \]

will generate a probabilistic interpretation. In this section, we will examine the probabilistic interpretation generated by

\[ G_{\text{4}}(x) = (e^x - 1) / (e^x + 1) \quad (20) \]

It is easy to show that this choice for \( G \) satisfies (19). Combining (16) - (20), we get

\[ \Delta_{\text{4}}(H,E,e) = \frac{p(H|E) - p(H|e)}{p(H|E) + p(H|e) - 2p(H|E)p(H|e)} \quad (21) \]

or, more simply,

\[ \Delta_{\text{4}}(H,E) = \frac{p(H|E) - p(H)}{p(H|E) + p(H) - 2p(H|E)p(H)} \quad (22) \]

As mentioned earlier, we denote a particular probabilistic interpretation with a subscript. We also need to subscript the function \( G \) since any function satisfying (19) produces a probabilistic interpretation. There is no need to subscript \( \Delta^{\infty} \) because \( F \) is fixed.
The parallel combination function for $\Delta_1$ can be derived from the combination function for $\Delta^\infty$. Recall that $\Delta^\infty$ was constructed such that it's parallel combination function is simply addition. That is,

$$\Delta^\infty(H, E_1, E_2) = \Delta^\infty(H, E_1) \cdot \Delta^\infty(H, E_2)$$  \hspace{1cm} (26)

This is also clear from (16). In combination with (17), this gives

$$C_1^{-1}(\Delta_1(H, E_1E_2)) = C_1^{-1}(\Delta_1(H, E_1)) \cdot C_1^{-1}(\Delta_1(H, E_2))$$ \hspace{1cm} (24)

and since

$$C_1^{-1} = \ln(1+x/1-x)$$ \hspace{1cm} (25)

we get

$$\frac{1 + \Delta_1(H, E_1E_2)}{1 - \Delta_1(H, E_1E_2)} = \frac{1 + \Delta_1(H, E_1)}{1 - \Delta_1(H, E_1)} \cdot \frac{1 + \Delta_1(H, E_2)}{1 - \Delta_1(H, E_2)}$$ \hspace{1cm} (26)

It is straightforward to show that $\Delta_1$ with the combination function (26) satisfies the basic axioms and the axioms of parallel combination as well as the requirements of a probabilistic interpretation. We say that $\Delta_1$ is a probabilistic interpretation for certainty factors.

The numerator of $\Delta_1$ clearly demonstrates that it is a probability update. In fact, $\Delta_1$ is not too different from the original definition of certainty factors. The numerators are the same while $\Delta_1$ seems to have a more complicated denominator. To get a better feel for $\Delta_1$, we can rewrite (22) as

$$\Delta_1(H,E) = \frac{p(H|E) - p(H)}{p(H)(1 - p(H))},$$ \hspace{1cm} (27)

Figure 3 depicts $\Delta_{orig}$ and $\Delta_1$ geometrically. Note the symmetry in the denominator of $\Delta_1$ that is lacking in the original interpretation.

$$\begin{align*}
\Delta_{orig} & = \frac{p(H|E) - p(H)}{p(H)(1 - p(H))} \\
\Delta_1 & = \frac{p(H|E) - p(H)}{p(E|He)p(H|He)}
\end{align*}$$ \hspace{1cm} (28)

A comparison of the combination functions

It is interesting to compare the parallel combination function we have derived for $\Delta_1$, (26), with the combination function used in MYCIN, (1). Although both functions obey the axioms of parallel combination, there is no reason to expect that the two functions are equivalent. It turns out, however, that they are close. In figure 4, the two functions are plotted for $CF(H,E_1) = \Delta_1(H,E_1) = 0.6$. The two curves never differ by more than a few percent. It can easily be shown that the two combination functions differ the most when

$$CF(H,E_1) = \Delta_1(H,E_1) \text{ approaches } 1,$$

and

$$CF(H,E_1) = \Delta_1(H,E_1) = -1 \cdot \left(1 - CF(H,E_1)\right)/2.$$ 

In this case, $CF(H,E,E_2)$ approaches -1/2 while $\Delta(H,E,E_2)$ approaches -1/3.

Figure 4: A comparison of the parallel combination function used in MYCIN and the parallel combination function for $\Delta_1$

Relationship to Bayes' theorem

In this section, we will show that the probabilistic interpretation $\Delta_1$ is consistent with Bayes' theorem under the assumption of conditional independence. To begin, consider Bayes' theorem for updating the probability of a hypothesis $H$ given prior evidence $e$ and new evidence $E$:

$$p(H|e) = \frac{p(E|H)e}{p(E|e)}.$$ \hspace{1cm} (29)

The corresponding formula for the negation of the hypothesis, $\overline{H}$, is

$$p(\overline{H}|e) = \frac{p(E|\overline{H})e}{p(E|e)}.$$ \hspace{1cm} (30)

Dividing (28) by (29) we get

$$p(H|e) \cdot \frac{p(E|\overline{H})}{p(E|H)} = \frac{p(E|H)e}{p(E|H)e} \cdot \frac{p(H|e)}{p(\overline{H}|e)}.$$ \hspace{1cm} (31)

Now the odds of some event $x$, denoted $O(x)$, is just

$$O(x) = \frac{p(x)}{1 - p(x)}$$

so that (30) can be written as

$$p(H|e) = \frac{p(E|H)e}{p(E|\overline{H})e}.$$ \hspace{1cm} (32)

Note that this form of Bayes' theorem was derived without assuming conditional independence of evidence.

Given (16) and (18) we can write

$$\Delta^\infty(H,E,e) = \ln\left(\frac{p(E|H)e}{p(E|\overline{H})e}\right)$$ \hspace{1cm} (33)

or using (31).

$$\Delta^\infty(H,E,e) = \ln\left(p(E|H)\cdot p(E|\overline{H})\right).$$ \hspace{1cm} (34)
Furthermore, with $G_1$ given by (20) we see that
\[ \Delta_1(H,E,e) = \frac{\mathbb{P}(E|H)e}{\mathbb{P}(E|\neg H)} - \frac{1}{\mathbb{P}(E|\neg H)e} + 1 \] (34)

or
\[ \frac{\mathbb{P}(E|H)e}{\mathbb{P}(E|\neg H)} = \left( \frac{\Delta_1(H,E,e) - 1}{\Delta_1(H,E,e) - 1} \right) \frac{\mathbb{P}(E|\neg H)}{\mathbb{P}(E|H)e} + 1 \] (35)

Now assume that $E$ and $e$ are conditionally independent given $H$ and $\neg H$. That is
\[ \mathbb{P}(E|H)e = \mathbb{P}(E|\neg H)e = \mathbb{P}(E|\neg H)e \] (36)

for all $e$ which does not entail $E$. Given this assumption, we see that the modularity axiom for $\Delta_1$, (4), is satisfied. Also, under assumption (36) we can write Bayes' theorem for two pieces of evidence as
\[
\begin{align*}
\mathbb{P}(E_1|E_2) & \quad \mathbb{P}(E_2|H)\mathbb{P}(H)\mathbb{P}(E_1|H)\mathbb{P}(H) \\
\mathbb{P}(E_1|\neg H) & \quad \mathbb{P}(E_2|H)\mathbb{P}(H)\mathbb{P}(E_1|\neg H)\mathbb{P}(H) \\
\mathbb{P}(E_2|E_1) & \quad \mathbb{P}(E_1|H)\mathbb{P}(H)\mathbb{P}(E_2|H)\mathbb{P}(H) \\
\mathbb{P}(E_2|\neg H) & \quad \mathbb{P}(E_1|H)\mathbb{P}(H)\mathbb{P}(E_2|\neg H)\mathbb{P}(H) \\
\end{align*}
\]

Substituting (35) into (37) then gives the parallel combination function for $\Delta_1$, (26).

Therefore, the probabilistic interpretation $\Delta_1$ for certainty factors is consistent with Bayes' theorem with the assumption of conditional independence of evidence on $H$ and its negation. Later we will see that the axioms of certainty factors and the requirements of a probabilistic interpretation for certainty factors mandates the conditional independence assumption.

**Relationship to the likelihood ratio**

The quantity $\mathbb{P}(E|H)/\mathbb{P}(E|\neg H)$ is commonly called the **likelihood ratio** and written $\lambda(H,E)$. With this shorthand, we can write Bayes' theorem for two pieces of evidence as
\[
\mathbb{P}(E_1, E_2) = \mathbb{P}(E_1|H)\mathbb{P}(E_2|H)\mathbb{P}(H)\mathbb{P}(E_1|\neg H)\mathbb{P}(E_2|\neg H)\mathbb{P}(H) \\
\mathbb{P}(E_2, E_1) = \mathbb{P}(E_1|H)\mathbb{P}(E_2|H)\mathbb{P}(H)\mathbb{P}(E_1|\neg H)\mathbb{P}(E_2|\neg H)\mathbb{P}(H) \\
\mathbb{P}(E_1) = \mathbb{P}(E_1|H)\mathbb{P}(H)\mathbb{P}(E_1|\neg H)\mathbb{P}(H) \\
\mathbb{P}(E_2) = \mathbb{P}(E_2|H)\mathbb{P}(H)\mathbb{P}(E_2|\neg H)\mathbb{P}(H) \\
\mathbb{P}(H) = \mathbb{P}(H) \\
\mathbb{P}(\neg H) = \mathbb{P}(\neg H) \\
\]

This equation is often called the **odds-likelihood form** of Bayes' theorem. Under the assumption of conditional independence, the parallel combination function for $\lambda(H,E)$ is simple multiplication:
\[ \lambda(H,E) = \lambda(H)\lambda(E) \] (39)

This follows from the definition of lambda, (38) and (37). The parallel combination function for the logarithm of the likelihood ratio, $\ln(\lambda(H,E))$, which is identical to $\Delta_1(H,E)$, is just addition. That is,
\[ \ln(\lambda(H,E)) = \ln(\lambda(H)) + \ln(\lambda(E)) \] (40)

This desirable property of the log-likelihood ratio was first noted by Peirce in 1870. He called the quantity $\ln(\lambda(H,E))$ the "weight of evidence" for a hypothesis and argued that it is a natural form for the expression and processing of beliefs. This property of the log-likelihood ratio was also discovered independently by Turing, Good, and Minsky. The additive property of the log-likelihood ratio also reveals a striking resemblance to the ad hoc scoring scheme of INTERNIST-I. Indeed, it is not surprising that such a simple and powerful method for reasoning with uncertainty has been "discovered" many times.

The likelihood ratio is used explicitly in several expert systems. For example, the GLASSOW DYSPEPSIA system uses $\ln(\lambda(H,E))$ and PROSPECTOR uses $\lambda(H,E)^4$. PROSPECTOR is of special interest because Duda, one of its designers, noticed a similarity between $\lambda(H,E)$ and certainty factors. In particular, he showed that the combination function for the quantity $\lambda(H,E)^4 - 1/\lambda(H,E^4)$, which he called $C(H,E)$, is quite similar to the MYCIN parallel combination function (1). From equation (34), we see that $C(H,E)$ is equal to $\Delta_1(H,E)$ and so the combination function for $C$ is just (26). Duda considered the combination function for $C$ to be an approximation of the combination function for CF. The concepts developed in this paper suggest that the one is no "better" than the other; both satisfy the axioms of parallel combination.

**Other probabilistic interpretations**

The mapping scheme developed above can be used to create an infinite number of valid interpretations. Recall that any $G$ which satisfies (19) will generate a probabilistic interpretation. One legitimate choice for $G$ is of special interest. It is easy to show that the map
\[
G(x) = \begin{cases} 
\ln(1/x) & x > 0 \\
\ln(x) & x < 0 
\end{cases}
\]

corresponding to the probabilistic interpretation
\[
\Delta_{CF}(H,E) = \begin{cases} 
\mathbb{P}(H|E) - \mathbb{P}(H) & \mathbb{P}(H|E) > \mathbb{P}(H) \\
\mathbb{P}(H|E)(1 - \mathbb{P}(H)) & \mathbb{P}(H|E) = \mathbb{P}(H) \\
\mathbb{P}(H)(1 - \mathbb{P}(H|E)) & \mathbb{P}(H|E) < \mathbb{P}(H) 
\end{cases}
\]
gives the MYCIN parallel combination function exactly. That is, there is a probabilistic interpretation for certainty factors which yields the combination function used by MYCIN. We will not dwell on this mapping since we have shown that differences between the $\Delta_1$ and MYCIN parallel combination functions are small. Also, the latter mapping is more difficult to analyze and work with.

Consider the following set of choices for $G$.
\[
G_n(x) = \begin{cases} 
e^{-1/n} & n = 1, 2, 3, \ldots \\
e^{-1/n} + 1 & x \leq 0 
\end{cases}
\]

It is easy to check that each $G_n$ produces a valid probabilistic interpretation
\[ \Delta_n(H,E) = \begin{cases} 
(\mathbb{P}(H|E)/\mathbb{P}(H))^{1/n} - 1 & \mathbb{P}(H|E) > \mathbb{P}(H) \\
(\mathbb{P}(H|E)/\mathbb{P}(H))^{1/n} + 1 & \mathbb{P}(H|E) = \mathbb{P}(H) 
\end{cases} \]

The parallel combination function corresponding to $G_n$ is given by
\[ \Delta_n(H,E) = \left[ \ln(1 - \Delta_n(H,E)) \right]^{1/n} + 1 \]

As $n$ approaches infinity, the combination function tends to select the certainty factor with the largest absolute value. Thus, the combination function can take on different qualitative behaviors depending on the probabilistic interpretation. In the next section, we will examine an

---

*This was first noted in a paper by Hajek.\(^2\)
*This example is given in a paper by Spiegelhalter.\(^8\)
*We will see later that $G$ affects the sequential combination function as well.
important implication of this fact. Here, we mention that (42) provides an explanation for a concern that the creators of certainty factors had about the behavior of their parallel combination function (1). They felt that combinations of positive evidence approached certainty (CF=1) too quickly. Perhaps they had the probabilistic interpretation \( \Delta \) (a large) in mind.

**Eliciting certainty factors from experts**

Certainty factors are often elicited from experts with rather vague phrases like: "Given E how strongly do you believe in H?" or "How strongly does E confirm H?" For example, these phrases were used during the construction of early versions of the MYCIN knowledge base.24 Undoubtedly, the lack of a clear operational meaning for certainty factors is responsible for the use of imprecise phrases such as these. Moreover, such imprecision is troubling in light of the fact that different probabilistic interpretations can dictate parallel combination functions with qualitatively different behaviors. However, once certainty factors are given a probabilistic interpretation, it is a straightforward task to elicit them precisely.

Equations (32) and (33) show that the ratio \( \text{O}(H|E)/\text{O}(H) \) or the equivalent ratio \( p(H|E)/p(E|H) \) is sufficient to determine \( \Delta \) for any probabilistic interpretation. This suggests two methods for eliciting certainty factors. The first is to elicit \( \text{O}(E|H) \) and \( p(E|H) \) separately and then use (34) to calculate \( \Delta(H,E) \). The second is to elicit \( p(E|H) \) assuming some \( p(H) \) and then use (22). The second method has an interesting special case. If we let \( p(H) = 1/2 \), (22) gives

\[
\Delta(H,E) = 2p(H|E) - 1
\]  

That is, suppose we ask the expert to imagine that the prior probability of some hypothesis is 1/2. Then we ask him for the posterior probability of the hypothesis given a piece of evidence. (43) states that this task of "starting" at 1/2 and giving a posterior probability is equivalent to the task of "starting" at 0 and giving a certainty factor in \([-1,1]\). This is shown diagrammatically in figure 6.

![Figure 6: A correspondence between eliciting probabilities and eliciting certainty factors](image)

It should be mentioned that eliciting probabilities is not a trivial task. Errors of bias are a major problem. For example, people tend to anchor on prior probabilities and tend to be conservative near 0 and 1 (i.e., when probabilities are near one, they are underestimated and when probabilities are near zero, they are overestimated). There are techniques for overcoming biases but such a discussion is not appropriate here.

**Completeness of the mapping scheme**

Earlier, we saw that it is possible to generate an infinite number of probabilistic interpretations for certainty factors with the map F given by (16) and the map G defined by (17). In this section, we show that this mapping scheme can generate all probabilistic interpretations. In other words, given any quantity which satisfies the basic axioms of certainty factors and the axioms of parallel combination as well as the requirements for a probabilistic interpretation, there is an F and G which generates it.

Consider the second half of the mapping scheme, (17). It turns out that certainty factors, as defined by the axioms, form a mathematical structure called an ordered commutative group. This was first noted by Hajek.12 A consequence of this fact is that given any quantity \( \Delta \) and associated parallel combination function which satisfies the axioms, there is a function \( \gamma \) such that

\[
\gamma(\Delta(H,E_1E_2)) = \gamma(\Delta(H,E_1)) \cdot \gamma(\Delta(H,E_2))
\]

where \( \gamma \) must satisfy

\[
\gamma(x) = -\gamma(-x)
\]

\[\gamma(1) = \infty\]

In other words, \( \Delta(H,E_1) \) and \( \Delta(H,E_2) \) can combined by first mapping them to another space with the function \( \gamma \), adding them in this new space, and then mapping the sum back to the original space with \( \gamma^{-1} \). However, \( \gamma^{-1} \) corresponds to what we have been calling G and \( \gamma(\Delta(H,E)) \) is just \( \Delta^{\infty}(H,E) \). Therefore, given a quantity which satisfies the axioms of certainty factors, we will be able to find a G such that (17) and (23) are satisfied. The requirements on choices for G, (19), follow directly from (44).

Now consider the first half of the mapping scheme, (16). First note that the function G has an inverse because it is monotonic. Therefore, we can rewrite the first requirement for a probabilistic interpretation, (12), in terms of \( \Delta^{\infty} \). That is, we require that there is some function \( f \) such that

\[
\Delta^{\infty}(H,E,E) = f(p(H|E,E), p(H|E))
\]  

But (45) and (23) together imply that

\[
f(p(H|E,E), p(H)) = f(p(H|E,E), p(H|E), p(H|E)) \cdot f(p(H|E), p(H))
\]

In other words, we know \( f \) satisfies

\[
f(x,z) = f(x,y) \cdot f(y,z)
\]

for all \( x, y, z \) between 0 and 1. Since \( f \) is smooth, there is some function \( \phi \) such that

\[
f(x,y) = \phi(x) \cdot \phi(y)
\]

However, it is clear from (16) and (45) that the function \( \phi \) is just what we have been calling F. In other words, the relationship between \( p(E|H) \) and \( p(H) \) and \( \Delta^{\infty} \) must be given by (16). Combining this result with the above we see that the mapping scheme, (16) and (17), can generate all probabilistic interpretations.

**The assumption of conditional independence**

In the appendix, constraints for the function F similar to those for G in the previous section are derived. These constraints are intimately related to the assumption of conditional independence, (36). In fact, is shown that if the desiderata of parallel combination are to be respected, the only reasonable assumption to make about the dependencies among pieces of evidence which bear on a hypothesis is that of conditional independence.

In this section, we examine two important restrictions placed on the use of certainty factors by the assumption of conditional independence.* The first restriction is that certainty factors cannot be used to manage uncertainty in an inference net when any set of mutually exclusive and exhaustive hypotheses in the net has more than two elements. To see this, suppose \( H_0, H_2, \) and \( H_3 \) are mutually exclusive and exhaustive. A problem arises when we assume that evidence is conditionally independent on the \( H_i \)’s and their negation. For example,  

*In a forthcoming paper, we will show that certainty factors can be generalized in such a way that the highly restrictive assumptions discussed in this section can be avoided.
We assume shortly. The expression which follows from the definition or conditional probability.

\[ p(E' \mid H, E) = p(E' \mid H) \]

And we get

\[ p(E \mid H) = p(E \mid H, E') = p(E \mid H, E) \]

which is not true in general.

The second restriction concerns two other types of combination defined in the certainty factor model. Those familiar with MYCIN may recall that the system uses “fuzzy set operators” to handle conjunctions and disjunctions of evidence in a rule premise. It is not very difficult to show that these functions violate the assumption of conditional independence. Therefore, these combinations must be avoided in order to adhere to the desiderata.

**Sequential combination**

Now let us turn our attention to sequential combination. Recall, that one hypothesis can serve as evidence for another. Therefore, evidence may itself be uncertain. As mentioned earlier, the simplest case can be represented as follows

\[ \Delta(E, E') \rightarrow \Delta(H, E) \rightarrow \Delta(H, E') \]

where \( E' \) is certain and \( E \) and \( H \) may not be certain. We are interested in constructing the single rule

\[ E' \rightarrow \rightarrow \rightarrow H \]

with certainty factor \( \Delta(H, E') \).

What properties are required of sequential combination? Unfortunately, no desiderata concerning sequential combination were given in the original work so a set of axioms that seem simple and intuitively appealing are introduced.

The first three axioms concern the value of \( \Delta(H, E') \) when \( \Delta(E, E') \) takes on the special values -1, 0, and 1. We require that

\[ \Delta(E, E') = 1 \] then \( \Delta(H, E') = \Delta(H, E) \).

\[ \Delta(E, E') = -1 \] then \( \Delta(H, E') = \Delta(H, E) \) and \( \Delta(H, E') = 0 \) and (46)

In the first case, \( E' \) proves \( E \) with certainty, in the second, \( E' \) disproves \( E \) with certainty and in the third, \( E' \) says nothing about \( E \). In the second case, it is clear that another certainty factor, \( \Delta(H, \sim E) \), must be associated with every IF-THEN rule. That is, it is necessary to know the update for \( H \) when \( E \) is false. From (2) we see that in MYCIN, \( CF(H, E') \) is 0 whenever \( CF(E, E') = -1 \). Thus, MYCIN is implicitly assuming \( CF(H, \sim E) = 0 \) for every rule. We will return to this point shortly.

The fourth axiom states that \( \Delta(H, E') \) is a function of only the certainty factors \( \Delta(H, E), \Delta(H, \sim E) \) and \( \Delta(E, E') \). Informally, we want the net update to depend only on some combination of the individual rule updates. The final requirement of the sequential combination function is that \( \Delta(H, E') \) is monotonic with respect to \( \Delta(E, E') \).

Before doing a formal derivation, let us consider sequential updating on more intuitive grounds. Consider the following expression which follows from the definition of conditional probability.

\[ p(E' \mid H, E') = p(E' \mid H) \]

In order to satisfy the fourth axiom of sequential propagation, we assume

\[ p(E' \mid H, E') = p(E' \mid H) \]

In this case, (47) becomes

\[ p(E' \mid E') = p(E' \mid E) \cdot p(E' \mid H) \cdot p(E' \mid H') \]

This formula is represented graphically in figure 6.

**Figure 6: Sequential updating in terms of absolute probabilities**

Before we know \( E', H \) and \( E \) are at their prior values, \( p(H) \) and \( p(E) \), respectively. The evidence \( E' \) then updates the probability of \( E \) to \( p(E') \) which changes the probability of \( H \) to \( p(H \mid E') \) (see the arrows in the above figure). In other words, there is a relationship between the updates \( \Delta(H, E') \) and \( \Delta(E, E') \).

We now consider the formal derivation of this relationship. Instead of generating an expression for \( \Delta(H, E') \) for some particular probabilistic interpretation, it is convenient to derive a relationship between the likelihood ratios

\[ \lambda(H, E') = \frac{p(E' \mid H)}{p(E' \mid \sim H)} \]

\[ \lambda(H, E') = \frac{p(E' \mid H)}{p(E' \mid \sim H)} \]

\[ \lambda(E, E') = \frac{p(E' \mid E)}{p(E' \mid \sim E)} \]

The expression for any particular probabilistic interpretation can then be computed from the relation

\[ \Delta(H, E) = \lambda(H, E') \Delta(E, E') \]

We begin with two expansions analogous to (48).

\[ p(E' \mid H) = p(E' \mid E)p(E \mid H) \cdot p(E' \mid \sim E)p(E \mid \sim H) \]

\[ p(E' \mid H) = p(E' \mid E)p(E \mid H) \cdot p(E' \mid \sim E)p(E \mid \sim H) \]

Dividing the first equation by the second we get

\[ \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

Now we can divide both numerator and denominator by \( p(E' \mid \sim E) \) and use \( p(E \mid H) + p(E \mid \sim H) = 1 \) to get

\[ \lambda(H, E') = \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

\[ \lambda(H, E') = \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

\[ \lambda(H, E') = \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

\[ \lambda(H, E') = \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

\[ \lambda(H, E') = \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

\[ \lambda(H, E') = \frac{p(E' \mid E)p(E \mid H)}{p(E' \mid \sim E)p(E \mid \sim H)} \]

Note that (48) implies

\[ p(E) = p(E \mid H)p(E) + p(E \mid \sim H)p(E) \]

so that the two priors, \( p(H) \) and \( p(E) \), cannot be specified independently. This led to problems in PROSPECTOR because experts were forced to provide priors and often did so incoherently. Instead of requiring consistent priors, PROSPECTOR accommodated the inconsistencies by using a non-Bayesian updating scheme. In this paper, we derive expressions relating probability update and, in doing so, avoid these inconsistencies.
Next, we use (50) and (51) to express \( p(E|H) \) and \( p(E|\neg H) \) in terms of \( \lambda(H,E) \) and \( \lambda(H,\neg E) \):

\[
\lambda(H,E) = \lambda(E|H)p(E|H) + \lambda(E|\neg H)p(E|\neg H)
\]

\[
\lambda(E'|H) = \lambda(E'|H)p(E'|H) + \lambda(E'|\neg H)p(E'|\neg H)
\]

Combining these results with (55) we get the sequential combination function in terms of \( \lambda \)’s:

\[
\lambda(H,E') = \frac{\lambda(H,E') \lambda(H,E)(1 - \lambda(H,E')) - \lambda(H,E')(1 - \lambda(H,E))}{\lambda(H,E) - \lambda(H,E')}
\]

We can derive the sequential combination function for the probabilistic interpretation \( \Delta_1 \) by using (53) with \( G \) given by (20). The result is

\[
\Delta_1(H,E') = \frac{-2 \Delta_1(H,E) \Delta_1(H,E') \Delta_1(E,E')}{\Delta_1(H,E) - \Delta_1(H,E')} \Delta_1(E,E') \Delta_1(H,E) \Delta_1(H,E')
\]

It is easy to show that (57) satisfies the axioms of sequential combination.

The sequential combination function for \( \Delta_1 \) is plotted in figure 7 for \( \Delta_1(H,E) = .9 \) and \( \Delta_1(H,\neg E) = -.25 \) and \( -.9 \). MYCIN’s sequential combination function is also plotted for comparison. The plot reveals considerable disagreement between the two functions. When \( \Delta_1(H,\neg E) \) is near \( \Delta_1(H,E) \), the two functions are close for \( \Delta(E,E') \) greater than 0 but diverge as \( \Delta(E,E') \) becomes more negative. When \( \Delta(E,E') \) becomes more positive, therefore we see that the axioms of sequential combination accommodate probabilistic interpretations whereas the particular sequential combination function used in MYCIN cannot have a probabilistic interpretation. Since well-understood probabilistic quantities can satisfy the intuitively appealing axioms of sequential combination, there is no reason to retain the original CF combination function.

Finally, let us consider the differences between the MYCIN and probabilistic combination functions in the example given earlier in this paper concerning the person who likes parties and solo backpacking trips. From the above discussion, it should be clear that we need to specify a certainty factor for the link

"EXTROVERT -------> DOES SOCIAL WORK"

Suppose this certainty factor is \( -.4 \). In this case, if we know a person likes to go to parties and also likes solo backpacking trips, then the increase in belief that he does social work is \( \Delta_1 = 2 \). This is close to \( CF = .24 \) calculated earlier. In fact, the difference is due to parallel combination only since the sequential combination functions for CF and \( \Delta_1 \) are identical when \( \Delta_1(H,E) = \Delta_1(H,\neg E) \) and \( \Delta(E,E') \) is greater than 0. However, suppose we only know that a person likes solo backpacking trips. In this case, we get \( CF = 0 \) and \( \Delta_1 = -.2 \) for the change in belief that he does social work. We see that the MYCIN combination function is ignoring the information contained in the link pictured above.

**Discussion**

Our analysis raises two important questions.

*Given the correspondence between certainty factors and probabilities, why use certainty factors at all? Or more precisely, why retain the distinction between certainty factors and probabilities?*

One possible answer to this question is that the concept of certainty factors or more generally, the concept of a probability update is important with respect to the user-system interaction. For example, the concept of an update can be used to assist in explanation. If a user were interested in knowing why a particular hypothesis was being considered, an expert system could display all evidence relevant to the hypothesis and the updates associated with each piece of evidence. In fact, "weight of evidence," a particular type of update mentioned earlier in this paper, is used by the GLASGOW DYSPEPSIA expert system in this way to explain its reasoning to the user22.

Updates can also be used for entering data into an expert system. To illustrate this, consider an expert system for the diagnosis of liver disease which has in its inference net a node "patient drinks alcohol frequently" with possible values "true" and "false." Suppose the program requests that the physician enter a value for this node. The physician then presumably asks the patient whether or not he drinks frequently. Based on the patient’s answer and how he answers it, the physician may not just want to enter "true" or "false" but express his uncertainty about the status of the node. It seems appropriate for the physician to relay his uncertainty as an update rather than an absolute probability. A simple argument for this can be made when the inference net has a tree structure. First note that in a tree network, prior probabilities for all non-root nodes can be calculated from the prior of the root node. In particular, the prior probabilities of the leaf or data nodes are

There is a simple intuitive explanation for the sharp concavity of the sequential combination function for \( \Delta_1 \) as \( \Delta_1(H,\neg E) \) approaches 0. This situation corresponds to the case where \( p(H) \) is near \( p(H|\neg E) \). In this situation, we see from figure 8 that \( p(E) \) is near 0. Since \( E \) is highly unlikely, it takes a strong positive update on \( E \) to raise its probability to a value that will alter the probability of \( H \) significantly.
function of the certainty factors or updates assigned to the network by some expert. Therefore, even though a user may agree with some specified prior for the root node, his priors for data nodes may not be consistent with those of the system because he may not have the same information that is contained in the net. Thus, if the physician simply reports an absolute probability for the status of the data node after the patient reports that he does not drink frequently, the resulting updates in the net may be inaccurate due to the differences between the physician's and system's prior probability for the node. A more accurate propagation of uncertainty occurs if the physician enters an update for the data node which can then be propagated according to the prescriptions of this work. Note that the physician does not have to calculate an update "in his head." He may simply enter both his prior and posterior probabilities for the node from which the system can calculate the update.

Why is MYCIN so successful?

In this paper, we discussed several restrictions that a system must satisfy before the use of certainty factors is consistent with the desiderata. However, MYCIN violates some of these restrictions. For example, the system contains sets of mutually exclusive and exhaustive hypotheses with more than two elements. In addition, certainty factors were elicited without a clear operational definition. In fact, they were elicited with the notion that they roughly corresponded to positive predictive value (p(H(E))), an absolute quantity. However, the system performs as well as experts in the field. Perhaps detailed considerations of uncertainty is not critical to the system's performance. Indeed, a sensitivity analysis of MYCIN's knowledge base carried out by Cooper and Clancey revealed that the system's performance did not depend strongly on changes in rule certainty factors. It would be interesting to study the performance of other systems using certainty factors with respect to systematic violations of the restrictions outlined in this paper.

Summary and conclusions

A redefinition of certainty factors in terms of the original desiderata has been proposed. The axioms of certainty factors have been examined in detail and it has been emphasized that these axioms reflect the notion that certainty factors represent a belief update. It has been shown that this redefinition accommodates an infinite number of probabilistic interpretations for certainty factors. These interpretations, in turn, have provided useful insights into necessary and sufficient assumptions for propagating certainty factors through an inference net. In particular, it has been shown that evidence which bears on any hypothesis in the net must be conditionally independent on the hypothesis and its negation. The analysis has made it clear that certainty factors should be elicited from experts in a precise fashion and straightforward methods for doing this have been presented. Specific improvements in the methods for sequential combination have also been presented. It is hoped that this discussion has clarified much of the mystery surrounding certainty factors.

Acknowledgements

I would especially like to thank Eric Horvitz for many thought-provoking discussions. I would like to thank Ben Grosos for pointing out the correspondence between Δ1's and the PROSPECTOR scoring scheme. Finally, I would like to thank Curt Langlotz, Glenn Rennels, Mark Musen, Peter Cheeseman, and Ted Shorlifl for their helpful suggestions.

Appendix: the assumption of conditional independence

Before examining the assumption of conditional independence, we will derive some general constraints on the function F which follow directly from the basic and parallel combination axioms and and the requirements for a probabilistic interpretation. Using the definition of G, (17), and the restrictions on this function (19), the second requirement of a probabilistic interpretation (13) becomes

$$\Delta_{\infty}^\infty(H,E) < \Delta_{\infty}^\infty(H,E) \iff p(H|E) < p(H)$$

Similarly, the third requirement (14) becomes

$$\Delta_{\infty}^\infty(H,E) = \infty \iff \{p(H|E) < 1 \text{ and } p(H) > 0\}$$

These requirements along with the definition of F, (16), give the following restrictions on F:

F monotonic increasing

$$F(1) = \infty$$

$$F(0) = -\infty$$

We are now ready to examine the relationship between the assumption of conditional independence and further restrictions on the function F. We begin the analysis by showing that if the function F is given by (18), then all evidence which bears directly on a hypothesis is conditionally independent of the hypothesis and its negation. As we showed earlier, when F is given by (18), the equations (16) and (31) give

$$\Delta_{\infty}^\infty(H,E,e) = \ln(p(E|H)e)p(E|He))$$

(see (33)). Since the function G has an inverse, we can write the modularity axiom (4) as follows:

$$\Delta_{\infty}^\infty(H,E) = \Delta_{\infty}^\infty(H,E)$$

for all e which does not entail E. Using (33) and the fact that the log function has an inverse, we get

$$p(E|He) p(E|H)$$

$$\Delta_{\infty}^\infty(H,E) = \Delta_{\infty}^\infty(H,E)$$

for all e which does not entail E. Note that (58) is almost the requirement of conditional independence.

When we considered sequential combination, we saw that every rule in an inference net of the form IF E THEN H must be associated with a certainty factor Δ(H,~E), the change in belief of H given the absence of E, as well as a certainty factor Δ(H,E). Given this, we get the companion requirement of (58)

$$p(E|He) p(E|H)$$

for all e which does not entail ¬E. From (58) and (59) it follows that either

$$p(E|He) p(E|H)$$

for all e which does not entail E or its negation, or that

$$p(E|He) = p(E|H)$$

for all e which does not entail E or its negation. (60) implies

$$\Delta(H,E) = 0$$

which means that E has no affect on H. Therefore, when F is given by (18), evidence which has any bearing on a hypothesis must be conditionally independent of the hypothesis and its negation.
The converse is also true. That is, if we assume conditional independence, F is essentially given by (18). To see this, first note that since F is monotonic increasing, we can write
\[
\Delta^{0}(h,E,e) = \ln(f'((0|h(e)/1-p(h|e))) - \ln(f'((p(h|e)/1-p(h|e))) (62)
\]
\[
\Delta^{0}(h,E,e) = \ln(f'(0|h(e))/f'(0|h(e)))
\]
where \(f'\) is a monotonically increasing, smooth function. In conjunction with the axiom of modularity, (4), this gives
\[
f'(0|h(e))/f'(0|h(e)) = \frac{f'(0|h(e))}{f'(0|h(e))} (63)
\]
for all \(e\) which does not entail \(E\). Now suppose all evidence for some hypothesis is conditionally independent on the hypothesis and its negation. Using this assumption and (31) we get
\[
O(h|e)/O(h|e) = O(h|e)/O(h) (64)
\]
for all \(e\) which does not entail \(E\). (63) and (64) together imply that
\[
f'(x) = \alpha^{a} x^{a}, \alpha > 0
\]
which means
\[
f'(x) = \alpha \ln(x/1-x) \quad \alpha > 0
\]
Since the factor \(\alpha\) can be incorporated into the definition of \(G\), we see that the assumption of conditional independence amounts to choosing \(F\) to be the function given by (18).

From the above analysis we see that the dependencies among evidence given hypothesis is directly related to the choice for \(F\). In particular, if we assume that \(F\) is not given by (65), then we must assume that all evidence which bears directly on a hypothesis is conditionally dependent in some uniform way determined by the exact choice for \(F\). The assumption of uniform conditional dependence seems far less attractive than the assumption of conditional independence. Therefore, we add the conditional independence assumption to the requirements for a probabilistic interpretation keeping in mind that there is no other reasonable alternative.

As mentioned in the text, it was the intention of the creators of certainty factors to require "modularity" since this would facilitate the construction of the knowledge base. The above analysis makes it clear that "modularity" corresponds to conditional independence in the probabilistic sense.

References
1. Shortliffe, E.H., and Buchanan, B.G., "A model of inexact reasoning in medicine," Mathematical Biosciences, No. 23, 1975, pp. 351-379.
2. Cohen, Paul R., Heuristic Reasoning About Uncertainty: An Artificial Intelligence Approach, PhD dissertation, Computer Science Department, Stanford University, August 1983, Rep. No. STAN-CS-83-986.
3. Lindley, D.V., "Scoring rules and the inevitability of probability," J. Stat. Plann. Eval., Vol. 50, 1992, pp. 1-26.
4. Cheesman, P., "In defense of probability", To be presented at the 9th International Joint Conference on Artificial Intelligence
5. Pearl, J., "Reverend Bayes on inference engines: a distributed hierarchical approach," Proceedings Second Annual Conference on Artificial Intelligence, Pittsburgh, Pennsylvania, AAAI, 1982, pp. 139-136.
6. Spiegelhalter, D.J., "A statistical view of uncertainty in expert systems", Presented at Workshop on AI and Statistics, Bell Laboratories, April 11-12, 1985
7. Charniak, E., "The Bayesian basis of common sense medical diagnosis," Proceedings of the National Conference on Artificial Intelligence, AAAI, 1983, pp. 70-73.
8. Yu, V.L., Fagan, L.M., Wraith, S.M., et al., "Antimicrobial Selection By A Computer: A Blinded Evaluation By Infectious Disease Experts," Journal of the American Medical Association, Vol. 242, No. 12, 1979, pp. 1279-1282.
9. Adams, R.B., "Probabilistic reasoning and certainty factors," Mathematical Biosciences, No. 32, 1976, pp. 177-186.
10. Shortliffe, E.H., "Certainty factors - a re-evaluation", Unpublished internal memo
11. Duda, R.O., "Alternative forms of Bayes rule for diagnosis," Tech. report, SRI, 1975, Project file 3805
12. Hajek, P., "Combining function for certainty degrees in consulting systems," Tech. report, Mathematical Institute, Prague, 1984.
13. Horvitz, E.J., and Heckerman, D.E., "Confusion between belief updating and absolute belief in artificial intelligence research", To be presented at the AAAI/IEEE Workshop on Uncertainty and Probability in Artificial Intelligence
14. Duda, R., Hart, P., and Nilsson, N., "Subjective Bayesian methods for rule-based inference systems," Proceedings 1976 National Computer Conference, Vol. 45, AFIPS, 1976, pp. 1075-1082.
15. Pearl, J., "Fusion, propagation, and structuring in Bayesian networks", Presented at the Symposium on Complexity of Approximately Solved Problems, Columbia University
16. Shortliffe, E. H., "Personal communication".
17. Peirce, C.S., The probability of induction, Simon and Shuster, 1958, pp. 1341-1354.
18. Good, I.J., The Bayesian influence, or how to sweep subjectivism under the carpet, D. Reidel Publishing Company, 1976, pp. 125-174.
19. Good, I.J., Probability and the weighing of evidence, Hafners, New York, 1950.
20. Good, I.J., "Weight of evidence, corroborations, explanatory power, information and the utility of experiments," J. R. Statist. Soc. B, No. 22, 1960, pp. 319-331.
21. Minsky, M., and Selfridge, O.G., Learning in random nets, Butterworths, London, 1961, pp. 335-347.
22. Spiegelhalter, D.J., and Knill-Jones, R.P., "Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology," *J. R. Statist. Soc. A*, No. 147, 1984, pp. 35-77.

23. Grosof, B., "Evidential confirmation as transformed probability", To be presented at the AAAI/IEEE Workshop on Uncertainty and Probability in Artificial Intelligence

24. Fagan, L. "Personal communication".

25. Tversky, A., and Kahneman, D., "Judgement under uncertainty: heuristics and biases," *Science*, Vol. 185, 1974, pp. 1124-1131.

26. Edwards, W., et al., "Special issue on probabilistic inference," *IEEE Transactions on Human Factors In Electronics*, Vol. 7, No. 1, 1966.

27. Spetzler, C.S., Staell von Holstein, C.S., *Probability encoding in decision analysis*, Strategic Decisions Group, 1984, pp. 603-625.

28. Lindley, D.V., Tversky, A., and Brown, R.V., "On the reconciliation of probability assessments," *J. R. Statist. Soc. A*, 1979, pp. 140-180.

29. Buchanan, B.G., Shortliffe, E.H., *Uncertainty and evidential support*, Addison-Wesley, Menlo Park, 1984, ch. 10.