Reasoning with Assertions of High Conditional Probability: Entailment with Universal Near Surety

Donald Bamber
Space & Naval Warfare Systems Center, San Diego
bamber@spawar.navy.mil

I. R. Goodman
Space & Naval Warfare Systems Center, San Diego
goodman@spawar.navy.mil

Abstract

Rules having rare exceptions are best interpreted as assertions of high conditional probability. In other words, a rule *If* X *then* Y is interpreted as meaning that $\Pr(Y|X) \approx 1$. In this paper, such rules are regarded as statements about imprecise probabilities, and imprecise probabilities are identified with convex sets of precise probabilities. A general approach to reasoning with such rules, based on second-order probability, is advocated. Within this general approach, different reasoning methods are needed, with the selection of a specific method being dependent upon what knowledge is available about the relative tightness of the approximation $\Pr(Y|X) \approx 1$ across rules. A method of reasoning, entailment with universal near surety, is formulated for the case when *no* knowledge is available concerning these relative tightnesses. Finally, it is shown that reasoning via entailment with universal near surety is equivalent to carrying out a particular test on a directed graph.

Keywords. Conditional probability, second-order probability, Bayesian inference, nonmonotonic logic, rule-based systems, threshold knowledge, informant, robustness, directed graph.

1 Second-Order-Probability Logic

Broadly speaking, this paper deals with the problem of estimating underconstrained conditional probabilities. By this, we mean the problem of estimating a desired conditional probability when (a) we are given information consisting of either exact values, approximations, or bounds for certain other conditional probabilities and (b) the given information fails to constrain the desired conditional probability to a unique value. This problem may be regarded as a problem in imprecise probability for the reason that the given information is consistent with more than one probability measure. Other investigators, e.g., [4, 6, 26, 27, 37], have attacked this problem using a variety of approaches. We have chosen to take a Bayesian approach in which we adopt a prior distribution over the space of probability measures and then base our conclusions on the posterior distribution resulting from conditioning on the given information [8, 9, 10, 12, 23, 24, 25]. Typically this involves finding the posterior expectation of the desired conditional probability. Because the prior and posterior distributions specify probabilities of probabilities, we call our general approach Second-Order-Probability Logic.

Motivated by concern over how reasoning should be done within rule-based systems, this paper deals with the special problem of ascertaining whether a desired conditional probability is close to one given that certain other conditional probabilities are close to one. To handle this special problem, we superimpose a certain limiting operation on our basic second-order Bayesian approach.

2 Rule-Based Systems

A rule-based system consists of the following: (a) a rule base, (b) a fact base, and (c) an inference engine.

Rule base. The rule base consists of a finite set Γ of, say, m rules, denoted $A_i \Rightarrow B_i, i = 1, \ldots, m$, and expressed in English as *If* A_i *then* B_i.

The rules typically have occasional exceptions. But, presumably no one would assert the rule *If* A_i *then* B_i unless it were the case that, whenever A_i is true, it is highly likely that B_i is true. Hence, any rule *If* A_i *then* B_i will be interpreted here as an assertion that the conditional probability $\Pr(B_i|A_i)$ is close to one.[1]

Fact base. The fact base consists of a collection of facts that describe the current case or situation. In a meteorological rule-based system, for example,

[1] In addition to *If* A_i *then* B_i, there are many ways of expressing in English, at least approximately, the concept that $\Pr(B_i|A_i) \approx 1$. Some other ways are B_i *is highly probable given* A_i and *Nearly all* A_i*s are* B_i*s*.

the fact base contains facts that describe the current weather. Unlike the rule base which remains constant, the fact base is always changing with the situation.

Inference engine. The inference engine is an algorithm that is designed to answer queries put to the rule-based system. Thus, the user of the system may want to know whether the potential fact D holds true for the current case or situation and, so, the user presents the inference engine with the query *Is D true?*. The inference engine responds to the query with an answer that is either affirmative, negative, or noncommittal.

General Principle of Query Answering. How should the inference engine answer such queries? Suppose that the fact base contains the k facts $C_1, \ldots, C_k$. Let C denote the conjunction of these facts. Suppose, further, that the query *Is D true?* has been presented to the inference engine. This query should be answered affirmatively if and only if it is believed that $\Pr(D|C) \approx 1$.[2] In other words, in the context of the facts C, the query *Is D true?* is really equivalent to asking *Is* $\Pr(D|C)$ *close to one?*. Notice that this latter question is a question about probability and, thus, it must be answered on the basis of our knowledge about probability. That knowledge is contained in the rule base, which tells us that certain conditional probabilities are close to one. Hence, the inference engine should give an affirmative answer to the query *Is D true?* if and only if, based on the rule base Γ, there is *reason to believe* that $\Pr(D|C) \approx 1$, i.e., that the rule $C \Rightarrow D$ is true.

The General Principle of Query Answering is imprecise in that it does not specify what criteria should be employed for judging whether there is *reason to believe* that $C \Rightarrow D$ holds. It turns out that there are various criteria, to be discussed below, that we might justifiably adopt for that purpose.

2.1 Role of Imprecise Probabilities

Anyone who has been informed of the rule $A_i \Rightarrow B_i$ but who has not been informed of the exact value of $\Pr(B_i|A_i)$ has probabilistic knowledge that is inherently imprecise. Thus, this paper is concerned with the problem of reasoning about *imprecise* probabilities. However, much of this paper will be about *precise* probabilities because we identify imprecise probabilities with sets of precise probabilities.

[2] Ideally, we would like to have $\Pr(D|C) = 1$. However, given that the rules in the rule base can be in error on rare occasions, the most we reasonably hope for is that our conclusions based on those rules will only rarely be in error. Thus, we must be satisfied with $\Pr(D|C) \approx 1$.

3 Preliminaries

Let $\mathcal{L}$ be a propositional language constructed from the primitive propositions $S_1, \ldots, S_r$ and the connectives $\wedge$ (*and*), $\vee$ (*or*), $\neg$ (*not*), and $\rightarrow$ (*material conditional*). Let $\perp$ be an abbreviation for some arbitrarily chosen contradiction, such as $S_1 \wedge \neg S_1$. If $A, B \in \mathcal{L}$, let $A \models B$ mean that A entails B in propositional logic and let $A \not\models B$ mean that A does not entail B. An *atom* is a proposition of the form $T_1 \wedge \cdots \wedge T_r$ where each T_i is either S_i or $\neg S_i$.

Definition 1 *For any propositions $A, B \in \mathcal{L}$, $A \Rightarrow B$ is an* assertion of high conditional-probability *(abbreviated* HCP assertion*).*

An HCP assertion $A \Rightarrow B$ is a syntactic object that is to be interpreted as expressing that $\Pr(B|A)$ is close to one. Thus, HCP assertions are not to be interpreted as propositions, but as *rules*.

Definition 2 *If* Pr *is any probability function on $\mathcal{L}$ and if X and Y are propositions in $\mathcal{L}$, define*

$$\Pr(Y|X) = \sup\{p \in [0,1] : \Pr(X \wedge Y) = p\Pr(X).\} \tag{1}$$

The above supremum-based definition of $\Pr(Y|X)$ agrees with the usual definition when $\Pr(X) > 0$ but, when $\Pr(X) = 0$, it has the effect of setting $\Pr(Y|X) = 1$ rather than leaving it undefined.

There are two powerful motivations for using this supremum concept of conditional probability. First, it creates an analogy between reasoning with HCP assertions and reasoning in classical logic. In the latter, there is no inconsistency in asserting both *All unicorns are white* and *All unicorns are not white* because both statements are true if unicorns do not exist. Analogously, there is no inconsistency in asserting that both *Nearly all unicorns are white*, i.e., $\Pr(\textit{white}|\textit{unicorn}) \approx 1$, and *Nearly all unicorns are not white*, i.e., $\Pr(\neg\textit{white}|\textit{unicorn}) \approx 1$, because, under Definition 2, both statements are true if $\Pr(\textit{unicorn}) = 0$.

Second, using the supremum concept of conditional probability increases the expressiveness of the language of HCP assertions. Thus, it is possible to construct an HCP assertion that may be interpreted as meaning that $\Pr(B|A)$ is *exactly* one, rather then merely close to one. As recognized by Adams [3, Corollary M3.4] who employed the supremum concept of conditional probability in [3], the HCP assertion $A \wedge \neg B \Rightarrow \perp$ may be interpreted as meaning that $\Pr(B|A)$ is *exactly* one. To see this, recall that $A \wedge \neg B \Rightarrow \perp$ means that $\Pr(\perp|A \wedge \neg B)$ is close to one. But, under Definition 2, $\Pr(\bot|A \wedge \neg B)$ has the value one, if $\Pr(A \wedge \neg B) = 0$, and has the value zero otherwise. Thus, $\Pr(\bot|A \wedge \neg B)$ is close to one if and only if $\Pr(A \wedge \neg B) = 0$. But, under Definition 2, $\Pr(A \wedge \neg B) = 0$ if and only if $\Pr(B|A) = 1$.

Critics of the supremum concept of conditional probability may be mollified by the fact that all of the mathematics presented in this paper can be developed without making any reference to conditional probability whatsoever. Specifically, under Definition 2, for any $\epsilon > 0$, the following statements are equivalent:

$$\Pr(Y|X) \geq 1 - \epsilon. \tag{2}$$

$$\Pr(X \wedge Y) \geq (1 - \epsilon)\Pr(X). \tag{3}$$

So, whenever an expression of the form (2) appears in Definitions 3 and 9 below, one may substitute an expression of the form (3). Thus, in this paper, the use of conditional probability in general and the supremum concept in particular is not a necessity. However, use of the supremum concept is a considerable convenience as it simplifies the formulation of mathematical statements that would otherwise be awkward.

Models. The intuitive idea of *model* in logic is that it is a *fully specified* state of affairs. Such a fully specified state of affairs is a *model of a statement* if the statement is true in that state of affairs. In a logic whose subject matter is probability, a state of affairs has been fully specified when the probability of every proposition has been specified. This motivates the following definition.

Definition 3 *Any probability function on $\mathcal{L}$ is said to be a* model. *The set of all these models is denoted* MD. *For any $\epsilon > 0$, the* models of $\Gamma = \{A_1 \Rightarrow B_1, \ldots, A_m \Rightarrow B_m\}$ under closeness parameter ϵ *are:*

$$\mathrm{md}_\epsilon(\Gamma) = \{\pi \in \mathrm{MD} : \pi(B_i|A_i) \geq 1 - \epsilon, i = 1, \ldots, m\}. \tag{4}$$

Similarly, the models of $C \Rightarrow D$ under closeness parameter ζ *are:*

$$\mathrm{md}_\zeta(C \Rightarrow D) = \{\pi \in \mathrm{MD} : \pi(D|C) \geq 1 - \zeta\}.$$

3.1 Consistency

Definition 4 Γ *is* consistent *if and only if* $\mathrm{md}_\epsilon(\Gamma)$ *is not empty for every $\epsilon > 0$. Let* MD^+ *denote*

$$\{\pi \in \mathrm{MD} : \pi(X) > 0 \text{ for all } X \in \mathcal{L} \text{ unless } X \models \bot\}.$$

Then Γ *is* Z-consistent *if and only if* $\mathrm{md}_\epsilon(\Gamma) \cap \mathrm{MD}^+$ *is not empty for every $\epsilon > 0$.*

The reason for the choice of term *Z-consistent* is that it turns out that Z-consistency is equivalent to the notion of consistency used in Pearl's System Z [32]. Necessary and sufficient conditions for both consistency and Z-consistency will be presented later in Theorem 1.

4 Entailment with Surety: Adams' Logic of Conditionals

Recall the General Principle of Query Answering enunciated earlier. If the fact base contains facts whose conjunction is C and if the rule base contains the set of rules Γ, then the query *Is D true?* should be answered affirmatively if and only if Γ gives us *reason to believe* the HCP assertion $C \Rightarrow D$. It was mentioned that this *reason to believe* could take different forms depending on what criterion for belief was adopted.

Let us now consider one criterion for believing $C \Rightarrow D$ given the rules Γ. This is the criterion adopted by Ernest Adams [1, 2, 3, 4] in his logics of conditionals and of high probability. Put loosely, Adams' criterion is that, given $\Gamma = \{A_1 \Rightarrow B_1, \ldots, A_m \Rightarrow B_m\}$ we should believe $C \Rightarrow D$ if and only if the conditional probabilities $\Pr(B_1|A_1), \ldots, \Pr(B_m|A_m)$ all being close to one guarantees that the conditional probability $\Pr(D|C)$ will be close to one. Whenever this criterion is satisfied, we may say that Γ *entails* $C \Rightarrow D$ *with surety.*[3] A formal statement of this criterion is given in the following definition.

Definition 5 *The set of HCP assertions* Γ entails with surety *the HCP assertion $C \Rightarrow D$ if and only if, for every $\zeta > 0$, there exists an $\epsilon > 0$ such that*

$$\mathrm{md}_\epsilon(\Gamma) \subseteq \mathrm{md}_\zeta(C \Rightarrow D). \tag{5}$$

Entailment with surety is *monotonic*, meaning that, if Γ entails $C \Rightarrow D$ with surety and if $\Gamma \subseteq \Gamma^*$, then Γ^* also entails $C \Rightarrow D$.

It is no easy matter to *directly* apply Definition 5 in order to ascertain whether Γ entails $C \Rightarrow D$ with surety. To overcome this difficulty, Adams did two things. First, he constructed a set of rules of inference [1, Definition 6] having the property that, when using those rules of inference to derive conclusions from Γ, $C \Rightarrow D$ is a derivable conclusion if and only if Γ entails $C \Rightarrow D$ with surety. Second, he constructed a decision procedure for testing whether or not Γ entails $C \Rightarrow D$ with surety [3, Meta-metatheorem 3]. (For a decision procedure with markedly improved efficiency, see [7].)

[3] This is not Adams' terminology. Adams used the term *probabilistic entailment* (abbreviated *p-entailment*) for what is called here *entailment with surety.*

Entailment with surety is equivalent or nearly so to reasoning in a number of other systems: a set of inference rules known as System P [28], a more concise set of inference rules [9, Definition 2.10], the Boolean algebra known as Product-Space Conditional-Event Algebra [24], coherent probability appraisals [20], conditional objects [14], universal possibilistic consequence [14], big-stepped probabilities [15] based on [36], and ϵ-belief functions [16]. In addition, Schurz [35] has shown that a corrected version of the propositional part of Delgrande's conditional logic [19] is equivalent to an extended form of Adams' logic [3].

4.1 Suitability for Rule-Based Systems

It appears that the reasoning criterion embodied in Definition 5 is more strict than we would typically want to use in rule-based systems. Thus, typical rule-based systems allow chaining of inferences. For example, if the fact A and the rule $A \Rightarrow B$ are known, the fact B may be inferred. Then, if the rule $B \Rightarrow C$ is also known, the fact C may also be inferred.

Recall, from the General Principle of Query Answering, that inferring C under the above circumstances is justifiable only if the rules $A \Rightarrow B$ and $B \Rightarrow C$ give reason to believe the rule $A \Rightarrow C$. But, in Adams' logic, the rules $A \Rightarrow B$ and $B \Rightarrow C$ do not entail with surety the rule $A \Rightarrow C$. The reason for this is that the conditional probabilities $\Pr(B|A)$ and $\Pr(C|B)$ both being close to one does not guarantee that $\Pr(C|A)$ will be close to one. In fact, it is easy to construct examples in which $\Pr(B|A)$ and $\Pr(C|B)$ can be made arbitrarily close to one and yet $\Pr(C|A)$ is zero.

The above discussion may be summarized:

Example 1 $\{A \Rightarrow B, B \Rightarrow C\}$ does not entail $A \Rightarrow C$ with surety.

The above example shows that, if we were to apply Adams' logic in a rule-based system, we would, in effect, be outlawing a method (inference chaining) that (a) is commonly employed in rule-based systems and (b) seems intuitively correct.

Is our intuition that $\Pr(C|A)$ should be close to one when $\Pr(B|A)$ and $\Pr(C|B)$ are both close to one fundamentally wrong? No, it's not. It is true that there do exist probability functions having the property that $\Pr(C|A)$is far from one even though $\Pr(B|A)$ and $\Pr(C|B)$ are both close to one. However, to anticipate Example 5 below, such probability functions are *rare*. Thus, if we judge $\Pr(C|A)$ to be close to one whenever $\Pr(B|A)$ and $\Pr(C|B)$ are both close to one, we will nearly always be correct.

The reason that Adams' logic and typical rule-based systems have different attitudes about the legitimacy of inference chaining is that they use different standards of evidence for reaching conclusions. Whereas Adams' logic is *explicitly* designed to require a high standard of evidence, typical rule-based systems are *implicitly* designed to require a not-so-high standard.

5 Entailment with Near Surety

5.1 Rationale

Note that, in Adams' logic, in order to infer $C \Rightarrow D$ from Γ, Eq. 5 of Definition 5 requires that *every* model of Γ (under closeness parameter ϵ) be a model of $C \Rightarrow D$ (under closeness parameter ζ). In the authors' view, this is too high a standard of evidence for use in rule-based systems. Instead, rule-based systems should generally be based on a logic in which a conclusion is inferred from premises whenever *nearly every* model of the premises is a model of the conclusion. Thus, $C \Rightarrow D$ should be inferred from Γ whenever *nearly every* model of Γ is a model of $C \Rightarrow D$. With such a logic, conclusions would be highly likely to be true rather than certain to be true. But, as a payoff for adopting this more lax criterion for inference, the system would infer more conclusions than it otherwise could. (For some logics that have used this general approach to lowering the standard of evidence needed to reach conclusions, see [6, 29].)

The first author has designed a logic, the Near-Surety Logic of HCP Assertions [9], in which, in order for a conclusion to be inferred from premises, nearly every model of the premises must be a model of the conclusion. In order to rigorously define what was meant by *nearly every model*, a probability measure over the space of models MD was employed. So, because the models themselves are probability functions, the probability measure over models is second-order.

5.2 Definitions

Definition 6 *If the number of primitive propositions in the propositional language $\mathcal{L}$ is r, then the number of atoms in the language is 2^r, which we will denote by n. Let these atoms be denoted $at_1, \ldots, at_n$. For any vector $\theta = (\theta_1, \ldots, \theta_n) \in \mathbf{R}^n$ whose coordinates are all non-negative and sum to one, let π_θ denote the unique probability function on $\mathcal{L}$ such that $\pi_\theta(at_j) = \theta_j$ for $j = 1, \ldots, n$. Thus, θ is a parameter that indexes the probability functions on $\mathcal{L}$.*

This a standard beginning for doing Bayesian statistical analyses [31, Introduction]. After indexing a set of probability functions with a parameter vector, one adopts a prior distribution for the parameter vector.

In our case, we will adopt, for our prior, the uniform distribution over the simplex of θ-vectors.[4] It is only convenient and not essential that the prior be uniform. Substitution of any of a wide range of non-uniform priors would not affect the logic that will be developed below [9, Section 3.1.1].

Since there is a one-to-one correspondence between the set MD of all probability functions on $\mathcal{L}$ and the simplex of θ-vectors in $\mathbf{R}^n$, it will be convenient to identify each probability function π with its parameter vector and to use the same symbol π for both. Thus, a set of probability functions may be regarded here as a set of vectors. In particular, this identification makes the set of models $\mathrm{md}_\epsilon(\Gamma)$ into a compact convex polytope in $\mathbf{R}^n$ [9, Proposition 2.14]. Conversely, the uniform prior distribution over θ-vectors may be regarded as a second-order distribution over a collection of first-order probability functions on $\mathcal{L}$.

Definition 7 *Assuming that* Γ *is consistent, let* $\Upsilon_{\mathrm{md}_\epsilon(\Gamma)}$ *denote a probability measure on the Borel-measurable subsets of* MD *that has support on the polytope* $\mathrm{md}_\epsilon(\Gamma)$ *and that is uniformly distributed on* $\mathrm{md}_\epsilon(\Gamma)$.

In essence, given the uniform second-order prior distribution on MD, $\Upsilon_{\mathrm{md}_\epsilon(\Gamma)}$ is the posterior distribution on MD conditioned on the occurrence of the event $\mathrm{md}_\epsilon(\Gamma)$.

Let $\Pi_{\mathrm{md}_\epsilon(\Gamma)}$ denote a random probability function having second-order probability distribution $\Upsilon_{\mathrm{md}_\epsilon(\Gamma)}$. In other words, $\Pi_{\mathrm{md}_\epsilon(\Gamma)}$ is a probability function selected at random from among all those that are models of Γ under closeness parameter ϵ. Then, $\Upsilon_{\mathrm{md}_\epsilon(\Gamma)}[\mathrm{md}_\zeta(C \Rightarrow D)]$ is the second-order probability that $\Pi_{\mathrm{md}_\epsilon(\Gamma)}(D|C)$ is at least $1-\zeta$.

Definition 8 Γ entails with near surety $C \Rightarrow D$ *if either (*a*)* Γ *is consistent and*

$$(\forall\zeta > 0)\left[\left(\lim_{\epsilon\to 0} \Upsilon_{\mathrm{md}_\epsilon(\Gamma)}[\mathrm{md}_\zeta(C \Rightarrow D)]\right) = 1\right], \quad (6)$$

*or (*b*)* Γ *is inconsistent.*

In the above definition, the reason for defining Γ to entail $C \Rightarrow D$ with near surety whenever Γ is inconsistent is the following. If Γ is inconsistent, then, for small ϵ, there exist no models of Γ and, therefore, every model of Γ is a model of $C \Rightarrow D$. A necessary and sufficient condition for entailment with near surety will be presented later in Theorem 2.

Equation 6 of the above definition is equivalent to stating that the random variable $\Pi_{\mathrm{md}_\epsilon(\Gamma)}(D|C)$ converges in probability to one as ϵ goes to zero. In other words, (6) states that, as ϵ goes to zero, it becomes more and more certain that nearly all models π of Γ under closeness parameter ϵ have the property that $\pi(D|C) \geq 1-\zeta$, no matter how small $\zeta > 0$ is taken to be. However, (6) does not require that this certainty ever become absolute.

Example 2 For premises, take $\Gamma = \{A \Rightarrow B\}$. As a potential conclusion take the contrapositive $\neg B \Rightarrow \neg A$. We want to know whether the premise entails the conclusion with near surety. It can be shown that

$$\Upsilon_{\mathrm{md}_\epsilon(A\Rightarrow B)}[\mathrm{md}_\zeta(\neg B \Rightarrow \neg A)] = \zeta/[\zeta + (1-\zeta)\epsilon]. \quad (7)$$

Note that, for every $\zeta > 0$ no matter how small, the right side of (7) approaches one as ϵ goes to zero. In other words, for sufficiently small $\epsilon > 0$, *nearly every* model π having the property that (i) $\pi(B|A) \geq 1-\epsilon$ also has the property that (ii) $\pi(\neg A|\neg B) \geq 1-\zeta$. However, (7) also shows that, no matter how small $\epsilon > 0$ is, it never becomes certain that a model having property (i) will also have property (ii). This shows that $A \Rightarrow B$ entails $\neg B \Rightarrow \neg A$ with near surety, but not with surety. (For further discussion of this example, including its vector geometry, see [9, Section 2.6].)

Second-order expectation. Letting E denote second-order expectation, Equation 6 of Definition 8 is equivalent to:

$$\lim_{\epsilon\to 0} E[\Pi_{\mathrm{md}_\epsilon(\Gamma)}(D|C)] = 1.$$

This shows that testing for entailment with near surety is an example of the kind of posterior expectation problem mentioned in Section 1 that is the subject of the authors' line of research [8, 9, 10, 12, 23, 24, 25].

Entailment with near surety is equivalent or nearly so to reasoning in a number of other systems: rational-closure [30], System-Z [32], least specific possibilistic consequence [14], and least-commitment consequence based on ϵ-belief functions [16].

Unlike entailment with surety, entailment with near surety is not monotonic. For further probabilistic approaches to constructing nonmonotonic logics, see [21, 17, 33, 34, 6] plus further references in [9].

6 The Importance of Thresholds

6.1 Equal Thresholds

The general intent of the definition of entailment with near surety (Definition 8) is that, given the HCP assertions Γ, there is reason to believe the HCP assertion $C \Rightarrow D$ if nearly every model of Γ is also a model of $C \Rightarrow D$.

[4] Distributions over this simplex are also employed in the statistical study of proportion vectors [5].

Which models are to be considered models of Γ? In Definition 8, a probability function π is considered to be a model of Γ if and only if each $\pi(B_i|A_i)$ reaches or exceeds a threshold and the *same threshold* $1-\epsilon$ is used for each $\pi(B_i|A_i)$.

The single-informant analogy. Under what circumstances, would it make sense for an agent to use the same threshold for every premise $A_i \Rightarrow B_i$? It certainly would make sense if the set Γ of HCP assertions had been obtained from an informant in the following manner. The informant has either full or partial knowledge of the "true" model π_{true}.[5] In addition, the informant has some threshold for classifying a conditional probability as close to one. Thus, the informant will classify the conditional probability $\pi_{\text{true}}(Y|X)$ to be close to one if $\pi_{\text{true}}(Y|X)$ exceeds his/her threshold. If the informant knows that $\pi_{\text{true}}(Y|X)$ does exceed the threshold, then the informant *may* report to the agent that $X \Rightarrow Y$ is a true HCP assertion, but will not do so otherwise.

Suppose that the agent and the informant are no longer in communication and the agent wants to know whether $C \Rightarrow D$ is a correct HCP assertion in the sense that $\pi_{\text{true}}(D|C)$ is is close to one. (Assume that the agent has his/her own threshold for judging closeness to one and isn't concerned with whether that threshold is the same as the informant's threshold.)

The agent's *only* knowledge about the "true" model π_{true} is that, as judged by his/her informant, $\pi_{\text{true}}(B_i|A_i), i = 1, \ldots, m$, are all greater than some threshold that we may denote by $1-\epsilon$. Hence, the agent knows that π_{true} is a member of the set $\text{md}_\epsilon(\Gamma)$. The only information that the agent has about the value of ϵ is that ϵ is small. So, the agent tries out various values of ϵ. If, for all really small values of ϵ, nearly all models $\pi \in \text{md}_\epsilon(\Gamma)$ have the property that $\pi(D|C)$ exceeds the agent's personal threshold $1-\zeta$ for judging closeness to one, then the agent concludes that he/she has good reason to believe $C \Rightarrow D$.

In essence, the definition of entailment with near surety (Definition 8) embodies an idealized version of the just-described strategy of the agent.

Note that the above strategy is susceptible to error. It's possible that $\pi_{\text{true}}(D|C)$ is far from one even though $\pi(D|C)$ is close to one for nearly all $\pi \in \text{md}_\epsilon(\Gamma)$. However, the likelihood of error is small.

Precise vs. imprecise probability. In the above, the agent's probabilistic knowledge is inherently imprecise. As is frequently done in the study of imprecise probabilities [18, 37], this paper represents the agent's inherently imprecise probability knowledge by a convex set of precise probability functions, specifically the parametrized set $\text{md}_\epsilon(\Gamma)$.

[5] For our purposes, it doesn't matter whether π_{true} is an aleatory probability or whether it represents the belief of a third party, such as an expert whom the informant has interviewed.

6.2 Unequal Thresholds

Real-life sources of HCP assertions. In real life, we are likely to obtain HCP assertions from a variety of sources: personal experience, experiences of acquaintances, newspaper stories, opinions of experts, and statistical studies. It seems unlikely that the sources would all employ the same threshold for judging closeness to one.

The fact that the different sources may employ different thresholds is a point that should be taken into account when attempting to deduce new HCP assertions from old ones. This was shown in the study of *scaled* HCP assertions.

Scaled HCP assertions. Without going into the full and mathematically precise details [9, Section 3], suffice it to say that a *scaled* HCP assertion has the form $X \Rightarrow^k Y$, where k is a positive integer or ∞, and may be loosely interpreted as meaning that $\Pr(Y|X)$ is at least as large as closeness-to-one threshold of $1-O(\epsilon^k)$. Entailment with near surety of scaled HCP assertions was defined in a manner analogous to entailment with near surety for ordinary unscaled HCP assertions. It was shown [9, Section 4.3] that, given Z-consistent premises, entailment with near surety was equivalent to deduction within Goldszmidt and Pearl's System-Z^+ [22].

It turns out that with scaled HCP assertions, the values of the closeness-to-one thresholds affect what conclusions are entailed. An example will illustrate this.

Example 3 Suppose that A, B, and C as well as P and Q are all propositions such that (i) $C \models B \models A$, (ii) $A \not\models B \not\models C \not\models \bot$, (iii) $Q \models P$, (iv) $P \not\models Q \not\models \bot$, and (v) $A \wedge P \models \bot$. Note that, from these properties, it follows that C and Q contradict each other, i.e., $C \wedge Q \models \bot$. Let:

$$\begin{aligned} \Theta_1 &= \{A \Rightarrow^h \neg B, B \Rightarrow^j \neg C\}. \\ \Theta_2 &= \{P \Rightarrow^k \neg Q\}. \end{aligned}$$

In the following, let $\models_{\text{ns}}$ denote entailment with near surety. Then, using Theorem 3.15 of [9], it can be shown:

$$\begin{aligned} \Theta_1 &\models_{\text{ns}} (C \vee Q) \Rightarrow^{h+j} Q. \\ \Theta_2 &\models_{\text{ns}} (C \vee Q) \Rightarrow^k C. \end{aligned}$$

Thus, Θ_1 and Θ_2 oppose each other as regards whether Q or C should be expected when $C \vee Q$ is true. What happens when Θ_1 and Θ_2 both hold? The answer depends on the relative sizes of $h+j$ and k. Thus, when $h+j>k$,

$$\Theta_1 \cup \Theta_2 \models_{\mathrm{ns}} (C \vee Q) \Rightarrow^{h+j-k} Q. \tag{8}$$

But, when $h+j<k$,

$$\Theta_1 \cup \Theta_2 \models_{\mathrm{ns}} (C \vee Q) \Rightarrow^{k-h-j} C. \tag{9}$$

Thus, the relative sizes of the deviations from unity of the three thresholds $1-O(\epsilon^h)$, $1-O(\epsilon^j)$, and $1-O(\epsilon^k)$ determines which of the conclusions (8) or (9), if either, that we should reach. So, if we didn't know the values of h, j, and k, then we aren't justified in reaching either of the conclusions (8) or (9).

6.3 Impact of Assuming Equal Thresholds

Assuming that thresholds are equal when they aren't can lead to errors of reasoning.

Example 4 Consider Example 3 again, but with ordinary (unscaled) HCP assertions replacing the scaled HCP assertions. It can be shown that

$$\{A \Rightarrow \neg B, B \Rightarrow \neg C, P \Rightarrow \neg Q\} \models_{\mathrm{ns}} \quad (C \vee Q) \Rightarrow Q. \tag{10}$$

This result is based on the assumption that the informants who supplied the HCP assertions on the left side of (10) all employed the same threshold $1-\epsilon$. But, we know from Example 3 that, if that assumption is not correct, then the conclusion on the right side of (10) is not justified. This shows that entailment with near surety can yield inappropriate conclusions when we don't possess the threshold knowledge implicitly assumed in the definition of entailment with near surety (Definition 8).

7 Entailment with Universal Near Surety

Proofs of the results in this section will appear in [11].

7.1 Reasoning when Threshold Knowledge is Minimal

The no-informant-in-common analogy. Suppose that the agent has obtained the HCP assertions in Γ from different informants with no two HCP assertions being supplied by a common informant. Suppose, furthermore, that the agent has no idea how the different informants' thresholds compare with each other. In other words, each HCP assertion $A_i \Rightarrow B_i \in \Gamma$ has been obtained from an informant having a closeness-to-one threshold of $1-\epsilon_i$ and, although the agent knows that all the ϵ_is are small, he/she has no idea of their relative sizes.

How should the agent deduce new HCP assertions when he/she has minimal threshold knowledge as described above? The following definition of *entailment with universal near surety* was designed as an answer to that question. Because the agent has no idea of the relative sizes of the ϵ_is, the limit operation in the following definition allows each ϵ_i to approach zero at its own rate which may be faster or slower than the other ϵ_is.

7.2 Definition

Definition 9 *Define the models of* Γ *under closeness parameters* $\epsilon_1, \ldots, \epsilon_m > 0$ *to be:*

$$\mathrm{md}_{\epsilon_1,\ldots,\epsilon_m}(\Gamma) = \{\pi \in \mathrm{MD} : \pi(B_i|A_i) \geq 1-\epsilon_i,\ i=1,\ldots,m\}.$$

If this set of models is not empty, let $\Upsilon_{\mathrm{md}_{\epsilon_1,\ldots,\epsilon_m}(\Gamma)}$ *denote a probability measure on the Borel-measurable subsets of* MD *that has support on and is uniformly distributed over* $\mathrm{md}_{\epsilon_1,\ldots,\epsilon_m}(\Gamma)$. *Then* Γ entails $C \Rightarrow D$ with universal near surety *if either (*a*)* Γ *is consistent and, for all* $\zeta > 0$,

$$\lim_{\epsilon_1,\ldots,\epsilon_m \to 0} \Upsilon_{\mathrm{md}_{\epsilon_1,\ldots,\epsilon_m}(\Gamma)}[\mathrm{md}_\zeta(C \Rightarrow D)] = 1, \tag{11}$$

*or (*b*)* Γ *is inconsistent.*

Equation 11 may be paraphrased as follows. Define a *threshold function vector* to be an m-tuple $(e_1, \ldots, e_m)$ of positive valued functions defined on $(0,1)$ and such that each $e_i(\epsilon) \to 0$ as $\epsilon \to 0$. Then, Eq. 11 is equivalent to stating that, for every choice of threshold function vector $(e_1, \ldots, e_m)$,

$$\lim_{\epsilon \to 0} \Upsilon_{\mathrm{md}_{e_1(\epsilon),\ldots,e_m(\epsilon)}(\Gamma)}[\mathrm{md}_\zeta(C \Rightarrow D)] = 1.$$

This last equation explains the meaning of the word *universal* in the phrase *entailment with universal near surety.* The convergence to one is universal across all threshold function vectors.

Consequently, if $C \Rightarrow D$ is entailed with near universal surety by Γ, then the agent will be nearly sure that $\pi_{\mathrm{true}}(D|C)$ is close to one, no matter what the relative sizes of the deviations $\epsilon_1, \ldots, \epsilon_m$ from unity of the informants' thresholds.

Entailment with universal near surety is robust in the following sense. When the agent has more threshold knowledge than the minimal knowledge implicitly assumed in the above definition, the agent may be justified in reaching additional conclusions that are not entailed with universal near surety. However, his/her additional threshold knowledge will never cause him/her to reject any conclusion that is entailed with universal near surety. In particular, any conclusion that is entailed with universal near surety will also be entailed with near surety. It will be seen however, in Example 6 below, that the converse does not hold.

Like entailment with near surety, entailment with universal near surety is nonmonotonic.

7.3 Some Useful Definitions

Definition 10 *If ξ denotes the HCP assertion $X \Rightarrow Y$, let $\xi^{\rightarrow}$ denote the corresponding material conditional $X \rightarrow Y$.*[6] *If Ξ is a set of HCP assertions, let $\Xi^{\rightarrow}$ denote the set of material conditionals $\{\xi^{\rightarrow} : \xi \in \Xi\}$.*

Definition 11 *If $\Psi \subseteq \Gamma$, let the set of HCP assertions in Ψ having antecedents that $\Psi^{\rightarrow}$* "excludes" *be denoted* EX Ψ. *That is:*

$$\mathrm{EX}\,\Psi = \{A_i \Rightarrow B_i \ \in \ \Psi : \Psi^{\rightarrow} \models \neg A_i\}.$$

A subset Ψ of Γ is self-excluding *if* EX $\Psi = \Psi$. *Let a non-increasing sequence of finite sets $\Gamma_0, \Gamma_1, \Gamma_2, \ldots$ be constructed as follows. Let $\Gamma_0 = \Gamma$. For $k \geq 0$, let $\Gamma_{k+1} = \mathrm{EX}\,\Gamma_k$. Note that there must exist some integer K such that $\Gamma_k = \Gamma_K$ for all $k > K$. Let $\Gamma_{\mathrm{se}} = \Gamma_K$.*

In Pearl's terminology [32], EX Ψ consists of those members of Ψ that are not "tolerated" by Ψ.

Proposition 1 *Γ_{se} is self-excluding and, if $\Psi \subseteq \Gamma$ is self-excluding, then $\Psi \subseteq \Gamma_{\mathrm{se}}$.*

Terminology. (From [13].) A *directed graph* or *digraph* consists of an ordered pair (V, A) in which V is a finite set and A is a subset of $\{(x, y) : x, y \in V \text{ and } x \neq y\}$. The elements of V are called *vertices* and the elements of A are called *arcs*. A *path* from $a \in V$ to $b \in V$ is a finite sequence of $k \geq 1$ distinct vertices $v_1, \ldots, v_k$ such that $v_1 = a$, $v_k = b$, and for $j = 1, \ldots, k-1$ the ordered pair (v_j, v_{j+1}) is an arc.

Definition 12 *Let* $\mathrm{D}(\Gamma)$ *denote a directed graph constructed from Γ as follows. The vertices of* $\mathrm{D}(\Gamma)$ *consist of all the subsets of Γ. The arcs of* $\mathrm{D}(\Gamma)$ *consist of all ordered pairs (Ψ, Ξ) such that*

$$\Gamma \supseteq \Psi \supset \Xi \supseteq \mathrm{EX}\,\Psi.$$

Let K denote the smallest value of k such that $\Gamma_k = \Gamma_{\mathrm{se}}$. Then, the sequence $\Gamma_0, \ldots, \Gamma_K$ is the path of steepest descent *from Γ to Γ_{se}.*

Definition 13 *A set of HCP assertions Ξ* supports *an HCP assertion $X \Rightarrow Y$ if* (a) *$\Xi^{\rightarrow} \not\models \neg X$ and* (b) *$\Xi^{\rightarrow} \models X \rightarrow Y$.*

In other words, Ξ supports $X \Rightarrow Y$ if $\Xi^{\rightarrow}$ implies $X \rightarrow Y$ but does not imply $X \rightarrow \neg Y$.

7.4 Key Results

Theorem 1 below is based on Theorem 3.14 and Proposition 4.8 of [9]. Theorem 2 is based on Theorem 3.17 of [9]. Theorem 3, which is the main theorem of this paper, is new.

Theorem 1 (a) *The set of HCP assertions Γ is consistent if and only if the set of propositions $\Gamma^{\rightarrow}$ is consistent.* (b) *Γ is Z-consistent if and only if $\Gamma_{\mathrm{se}} = \emptyset$.*

Theorem 2 *Γ entails $C \Rightarrow D$ with near surety if and only if either* (a) *the path, in* $\mathrm{D}(\Gamma)$*, of steepest descent from Γ to Γ_{se} contains a vertex that supports $C \Rightarrow D$ or* (b) *${\Gamma_{\mathrm{se}}}^{\rightarrow} \models \neg C$.*

Theorem 3 *Γ entails $C \Rightarrow D$ with universal near surety if and only if either* (a) *every path in* $\mathrm{D}(\Gamma)$ *from Γ to Γ_{se} contains a vertex that supports $C \Rightarrow D$ or* (b) *${\Gamma_{\mathrm{se}}}^{\rightarrow} \models \neg C$.*

A current research question is whether there exists a way of testing for entailment with universal near surety that is more efficient than examining every path from Γ to Γ_{se}. This question can be answered affirmatively in the following very special case.

Corollary 1 *Suppose that $\Gamma^{\rightarrow} \not\models \neg C$. Then Γ entails $C \Rightarrow D$ with universal near surety if and only if $\Gamma^{\rightarrow} \models C \rightarrow D$.*

7.5 Examples

Example 5 Recall from Example 1 that, if $\Gamma = \{A \Rightarrow B, B \Rightarrow C\}$, then Γ does not entail $A \Rightarrow C$ *with surety*. However, assuming that $A \wedge B \wedge C \not\models \bot$, then Γ supports $A \Rightarrow C$ and, therefore, by Corollary 1, Γ entails $A \Rightarrow C$ *with universal near surety*.

Example 6 Consider Example 4 once again. Let Γ consist of γ_1, γ_2, and γ_3, which respectively denote $A \Rightarrow \neg B$, $B \Rightarrow \neg C$, and $P \Rightarrow \neg Q$. Also, let ρ denote $(C \vee Q) \Rightarrow Q$.

We know from (10) of Example 4 that Γ entails ρ with near surety. Now, we want to know whether Γ entails ρ with *universal* near surety.

[6] Recall that, unlike the HCP assertion $X \Rightarrow Y$, the material conditional $X \rightarrow Y$ is an ordinary proposition in $\mathcal{L}$ and is equivalent to $Y \vee \neg X$.

In D(Γ), the paths from Γ to $\Gamma_{\text{se}} = \emptyset$ are the following:

$$\begin{array}{l} \Gamma,\ \{\gamma_2\},\ \emptyset. \\ \Gamma,\ \{\gamma_1,\gamma_2\},\ \{\gamma_2\},\ \emptyset. \\ \Gamma,\ \{\gamma_2,\gamma_3\},\ \{\gamma_2\},\ \emptyset. \\ \Gamma,\ \{\gamma_2,\gamma_3\},\ \{\gamma_3\},\ \emptyset. \end{array}$$

The first of these paths is the path of steepest descent. It contains a vertex, $\{\gamma_2\}$, that supports ρ. This shows, as we already knew, that Γ entails ρ with near surety. Now, notice the last path. None of the vertices in this path support ρ and, furthermore, $\Gamma_{\text{se}}{}^{\rightarrow} \not\models \neg(C \vee Q)$. Therefore, Γ does *not* entail ρ with universal near surety.

8 Summary

Motivated by the need in rule-based systems for a reasoning method less stringent than entailment with surety, two methods of reasoning with imprecise probabilities, namely, entailment with near surety and entailment with universal near surety have been defined and characterized.

Acknowledgements

We are indebted to Ernest W. Adams, Moisés Goldszmidt, Hung T. Nguyen, and William C. Torrez for helpful discussions on the topic of this paper.

The mailing address for both authors is:
SPAWARSYSCEN D44215 (PL-BK)
53345 RYNE ROAD ROOM 222
SAN DIEGO CA 92152-7251
USA

References

[1] E. W. Adams. Probability and the logic of conditionals. In J. Hintikka & P. Suppes (eds.), *Aspects of Inductive Logic*. Pp. 265–316. North Holland Publishing, 1966.

[2] E. W. Adams. *The Logic of Conditionals*. D. Reidel, 1975.

[3] E. W. Adams. On the logic of high probability. *Journal of Philosophical Logic*, 15: 255–279, 1986.

[4] E. W. Adams. Four probability-preserving properties of inferences. *Journal of Philosophical Logic*, 25: 1–24, 1996.

[5] J. Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall, 1986.

[6] F. Bacchus, A. J. Grove, J. Y. Halpern, & D. Koller. From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, 87: 75–143, 1996.

[7] D. Bamber. Probabilistic entailment of conditionals by conditionals. *IEEE Transactions on Systems, Man, and Cybernetics*, 24: 1714–1723, 1994.

[8] D. Bamber. How probability theory can help us design rule-based systems. *Proceedings of the 1998 Command & Control Research & Technology Symposium*, Naval Postgraduate School, Monterey, CA, June 29 – July 1, 1998. Pp. 441–451, 1998.

[9] D. Bamber. Entailment with near surety of scaled assertions of high conditional probability. *Journal of Philosophical Logic*, 29: 1–74, 2000.

[10] D. Bamber & I. R. Goodman. New uses of second order probability techniques in estimating critical probabilities in Command & Control decision-making. *Proceedings of the 2000 Command & Control Research & Technology Symposium*. Naval Postgraduate School, Monterey, CA, June 26–28, 2000. <http://www.dodccrp.org/2000CCRTS/cd/html/pdf_papers/Track_4/124.pdf>.

[11] D. Bamber & I. R. Goodman. Robust reasoning with assertions of near-unity conditional probability when knowledge about near-unity thresholds is minimal. To be submitted for publication.

[12] D. Bamber, I. R. Goodman, & H. T. Nguyen. Extension of the concept of propositional deduction from classical logic to probability: An overview of probability selection approaches. *Information Sciences*, 2001, in press.

[13] J. Bang-Jensen & G. Gutin. *Digraphs: Theory, Algorithms and Applications*. Springer, 2000.

[14] S. Benferhat, D. Dubois & H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92: 259–276, 1997.

[15] S. Benferhat, D. Dubois & H. Prade. Possibilistic and standard probabilistic semantics of conditional knowledge bases. *Journal of Logic and Computation*, 9: 873–895, 1999.

[16] S. Benferhat, A. Saffiotti & P. Smets. Belief functions and default reasoning. *Artificial Intelligence*, 122: 1–69, 2000.

[17] R. A. Bourne & S. Parsons. Maximum entropy and variable strength defaults. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99).* Stockholm, July 31–August 6, 1999. Vol. 1, pp. 50–55, 1999.

[18] F. Cozman. *Introduction to the Theory of Sets of Probabilities.* <http://www.cs.cmu.edu/˜qbayes /Tutorial/>.

[19] J. P. Delgrande. A first-order conditional logic for prototypical properties. *Artificial Intelligence*, 33: 105–130, 1987.

[20] A. Gilio. Precise propagation of upper and lower probability bounds in System P. *Annals of Mathematics and Artificial Intelligence*, in press.

[21] M. Goldszmidt, P. Morris & J. Pearl. A maximum entropy approach to nonmonotonic reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15: 220–232, 1993.

[22] M. Goldszmidt & J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84: 57–112, 1996.

[23] I. R. Goodman. A decision aid for nodes in Command & Control Systems based on cognitive probability logic. *Proceedings of the 1999 Command & Control Research & Technology Symposium*. U. S. Naval War College, Newport, RI, June 29–July 1, 1999. Pp. 898–941, 1999.

[24] I. R. Goodman & H. T. Nguyen. Adams' high probability deduction and combination of information in the context of product probability conditional event algebra. *Proceedings of the International Conference on Multisource-Multisensor Information Fusion (FUSION-98).* Las Vegas, NV, July 6–9, 1998. Vol. 1, pp. 1–8, 1998.

[25] I. R. Goodman & H. T. Nguyen. Computational aspects of quantitative second order probability logic and fuzzy if-then rules: Part 1, Basic representations as integrals. *Proceedings of the Fifth Joint Conference on Information Sciences.* Atlantic City, NJ, February 27–March 3, 2000. Vol. 1, pp. 64–67, 2000.

[26] T. Hailperin. *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications.* Associated University Presses, 1996.

[27] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70: 939–952, 1982.

[28] S. Kraus, D. Lehmann & M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44: 167–297, 1990.

[29] H. E. Kyburg, Jr. Combinatorial semantics: Semantics for frequent validity. *Computational Intelligence*, 13: 215–257, 1997.

[30] D. Lehmann & M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55: 1-60, 1992.

[31] D. V. Lindley. *Bayesian Statistics, A Review.* Society for Industrial and Applied Mathematics, 1971.

[32] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh (ed.), *Theoretical Aspects of Reasoning about Knowledge. Proceedings of the Third Conference (TARK 1990).* Morgan Kaufmann, pp. 121-135, 1990.

[33] G. Schurz. Probabilistic justification of default reasoning. In B. Knebel & L. D. Dreschler-Fischer (eds.), *KI-94: Advances in Artificial Intelligence. Proceedings of the German Annual Conference on Artificial Intelligence.* Springer, pp. 248–259, 1994.

[34] G. Schurz. Probabilistic default logic based on irrelevance and relevance assumptions. In D. Gabbay, *et al.* (eds.) *Qualitative and Quantitative Practical Reasoning*. Lecture Notes in Artificial Intelligence, Vol. 1244. Springer, pp. 536–553, 1997.

[35] G. Schurz. Probabilistic semantics for Delgrande's conditional logic and a counterexample to his default logic. *Artificial Intelligence*, 102: 81–95.

[36] P. Snow. Diverse confidence levels in a probabilistic semantics for conditional logics. *Artificial Intelligence*, 112, 269–279, 1999.

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