NORTH-HOLLAND

Fuzzy Independence and Extended Conditional Probability

MARK E. HOFFMAN

Systems and Services Department, Orange Research Inc., Milford, CT 06460 and Department of Computer Science, Polytechnic University, Brooklyn, NY 11201

L. M. MANEVITZ*

Department of Computer Science, Polytechnic University, Brooklyn, NY 11201 and Department of Mathematics and Computer Science, Haifa University, Haifa, Israel

and

EDWARD K. WONG

Department of Computer Science, Polytechnic University, Brooklyn, NY 11201

ABSTRACT

In many applications, the use of Bayesian probability theory is problematical. Information needed to feasibility calculate is unavailable. There are different methodologies for dealing with this problem, e.g., maximal entropy and Dempster-Shafer Theory. If one can make independence assumptions, many of the problems disappear, and in fact, this is often the method of choice even when it is obviously incorrect. The notion of independence is a 0-1 concept, which implies that human guesses about its validity will not lead to robust systems. In this paper, we propose a fuzzy formulation of this concept. It should lend itself to probabilistic updating formulas by allowing heuristic estimation of the "degree of independence." We show how this can be applied to compute a new notion of conditional probability (we call this "extended conditional probability"). Given information, one typically has the choice of full conditioning (standard dependence) or ignoring the information (standard independence). We list some desiderata for the extension of this to allowing degree of conditioning. We then show how our formulation of degree of independence leads to a formula fulfilling these desiderata. After describing this formula, we show how this compares with other possible formulations of parameterized independence. In particular, we compare it to a linear interpolant, a higher power of a linear interpolant, and to a notion originally presented by Hummel and Manevitz [*Tenth Int. Joint Conf. on Artificial Intelligence*,

* Larry Manevitz was a Visiting Associate Professor at Polytechnic University during the preparation of this paper.

INFORMATION SCIENCES **90**, 137–156 (1996)
© Elsevier Science Inc. 1996
655 Avenue of the Americas, New York, NY 10010

0020-0255/96/$15.00
SSDI 0020-0255(95)00240-5

1987]. Interestingly, it turns out that a transformation of the Hummel-Manevitz method and our "fuzzy" method are close approximations of each other. Two examples illustrate how fuzzy independence and extended conditional probability might be applied. The first shows how linguistic probabilities result from treating fuzzy independence as a linguistic variable. The second is an industrial example of troubleshooting on the shop floor.

1. INTRODUCTION

Changing one's beliefs when new information becomes available is a common mode of human reasoning. It is observed in the deliberate gathering of pertinent evidence during medical diagnosis, or industrial troubleshooting. Less obvious, but equally as effective, beliefs are changed by strategically presented evidence in political persuasion, and promises of power, wealth, fame, and sexual prowess in marketing. One model of belief updating is based on probability theory, and, in particular, the use of Bayes Theorem. Conditional probabilities indicate the dependencies between new and existing evidence. Often these probabilities are unavailable or difficult to obtain. Mitigating the problem, independence may be assumed. Since independence is a 0-1 concept, when assumed incorrectly it leads to poor estimates of updated beliefs. Different methodologies have been developed to deal with this problem. For example, maximum entropy is a strategy complementing probability theory to model belief updating [7]. The ignorance produced by assuming independence is uniformly distributed over the constrained distribution. Alternatively, Dempster-Shafer Theory maintains ignorance in the frame of discernment, and uncertainty in sets of hypotheses. As new evidence becomes available, it is combined with existing evidence using Dempster's Rule [4, 9]. Neither method directly addresses independence beyond its binary nature, and both use fairly costly updating methods. Assuming independence of evidence simplifies probability estimation, problem formulation, and computation. On the other hand, requiring dependency demands detailed probability estimates and more complicated formulation and computation. Updating methodologies to work around the independence-assumption problem may ease the estimation problem, but at a higher cost in problem formulation and computation. In this paper, we present an approach that retains the benefits of the independence assumption for estimation, and includes dependency information in a less complex and more intuitive way using fuzzy independence.

A notion beyond that of binary independence does not appear to be well represented in the literature. Fine [3] reviews conditional probability

and independence from several theoretical standpoints. However, he does not present any notion of independence beyond that of binary independence. Subjective probabilists embody dependency information in conditional probabilities [2, 11]. In fact, they axiomize conditional probability as a personal degree of belief grounded in decision theory. Also, they give guidelines for estimating conditional probabilities and, thereby, estimate dependencies. Pearl's work on Bayesian Networks [11] embeds the subjective probabilistic approach in graphs. The notion of independence is central to the formulation; however, it remains binary in nature. The separate, but related, notion of *relevance* seems closer to the concept we are investigating, but it is not developed beyond a qualitative description. Hummel and Manevitz [5, 6] offer theoretical work directly addressing dependency beyond the concepts of 0-1 independence and conditional probability estimates. They approach the problem from a statistical standpoint updating information about the mean and covariance of probability distributions. They suggest an exponential formulation called *α-dependence*, and propose a method to estimate it.

In this work, we use fuzzy set theory [12] to extend a widely accepted probabilistic interpretation of set-theoretic events within a sample space [3, 10]. Fuzzy set theory extends the standard concept of set membership from discrete $\{0, 1\}$ to continuous $[0, 1]$. It allows us to vary the significance of one event with respect to another by continuously varying its contrast to its background. We develop the degree of independence concept, and use it to extend conditional probability. We show that extended conditional probability satisfies reasonable properties for both unconditional and conditional independence. A special case of extended conditional probability is developed, and approximations are offered to simplify computations. Finally, we offer two applications, one to linguistic probability and the other to industrial troubleshooting.

We conclude that the relaxation of independence in this way is a new concept that captures information that is known to exist, but is often ignored, in a direct and uncomplicated manner. Degree of independence is a vague and intuitive concept, and values are estimates; therefore, approximations that allow simple estimation are appropriate. Extended conditional probability retains the simplicity of assuming independence, and it incorporates dependency estimates in an intuitive way at a relatively small cost.

Section 2 presents conditional probability defined in terms of binary independence. In Section 3, properties are presented for the extension of conditional probability to include values between those designated by standard independence and full conditioning. The failure of a simple probabilistic formulation that satisfies the properties, and the success of a

fuzzy formulation, are presented in Section 4. Section 5 introduces a special case of extended conditional probability, and offers approximations that simplify estimation and computation. Applications to linguistic probability and industrial troubleshooting are shown in Section 6. Sections 7 and 8 offer discussion, conclusions, and topics for further research.

2. CONDITIONAL PROBABILITY AND BINARY INDEPENDENCE

Let A and B be two events in sample space Ω. If $P(A \cap B) = P(A) \cdot P(B)$, then events A and B are *independent*. Conditional probability is defined as $P(A|B) = P(A \cap B)/P(B)$. When events A and B are independent, $P(A|B) = P(A)$ [10].

If we define $u(A, B)$ as the independence of events A and B, then

$$u(A, B) = \begin{cases} 1, & \text{if } A \text{ and } B \text{ are independent,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

As defined, independence takes only two values, 0 or 1, depending upon whether events A and B are independent. This can be called *binary independence*. Subsequently, conditional probability can be expressed

$$P(A|B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)}, & \text{if } u(A, B) = 0, \\ P(A), & \text{if } u(A, B) = 1. \end{cases}$$

3. EXTENDED CONDITIONAL PROBABILITY AND ITS PROPERTIES

It can be observed that $P(A) = P(A \cap \Omega)/P(\Omega)$. That means that when $u(A, B) = 1$, then $P(A|B) = P(A \cap \Omega)/P(\Omega)$, and when $u(A, B) = 0$, then $P(A|B) = P(A \cap B)/P(B)$. It seems reasonable to extend conditional probability to include values between $P(A \cap B)/P(B)$ and $P(A \cap \Omega)/P(\Omega)$. Since these are the extremes of full conditioning and full independence, respectively, values for partial conditioning should be bounded by them. To include these values, we need to extend the concept of independence beyond that of binary independence; we allow values of independence to range between 0 and 1, inclusive. Informally, A and B can be 0.9 independent. That means that, given information about the

independence of A and B, a choice between full conditioning (standard dependence) and ignoring information (standard independence) does not have to be made. Using this amplified notion of independence, we extend conditional probability to P′(A|B), which we shall call *extended conditional probability* (ECP). $P'(A|B)$ should have the following desirable properties:

I. $P'(A|B)=P(A)$, if $u(A,B)=1$.

II. $P'(A|B)=P(A\cap B)/P(B)$, if $u(A,B)=0$.

III. $P'(A|B)$ is bounded by $P(A)$ and $P(A\cap B)/P(B)$.

IV-A. If $P(A)<P(A\cap B)/P(B)$, then $P'(A|B)$ is nonincreasing as the independence of A and B goes from 0 to 1.

IV-B. If $P(A)>P(A\cap B)/P(B)$, then $P'(A|B)$ is nondecreasing as the independence of A and B goes from 0 to 1.

IV-C. If $P(A)=P(A\cap B)/P(B)$, then $P'(A|B)=P(A)=P(A\cap B)/P(B)$ as the independence of A and B goes from 0 to 1.

Properties **I** and **II** guarantee that, for the extremes of full conditioning and full independence, ECP reduces to the standard definition of conditional probability, and Property **III** guarantees that ECP is bounded by these extremes. Properties **IV-A** and **IV-B** describe the monotonicity of ECP as the independence of A and B goes from 0 to 1 depending on which case has a larger probability. Property **IV-C** says that if the full-conditioning and full-independence cases are equal, ECP will be that value, also, regardless of the independence of A and B.

4. PROBABILISTIC AND FUZZY FORMULATIONS OF EXTENDED CONDITIONAL PROBABILITY

A simple probabilistic formulation of ECP would be based on increasing the size of B as the independence of A and B goes from 0 to 1. As the independence of A and B approaches 1, B approaches Ω. To expand B, elements may be added as full members; B may be expanded by any element in Ω not already in the expanded B. Let C be some set such that $B\subset C\subset\Omega$; it should be the case that

$$\frac{P(A\cap B)}{P(B)}>\frac{P(A\cap C)}{P(C)}>\frac{P(A\cap\Omega)}{P(\Omega)},\quad \text{if } \frac{P(A\cap B)}{P(B)}>P(A).$$

A similar statement can be made for "$<$." However, given only that $B\subset C\subset\Omega$, the above statement is not always true. A counterexample will illustrate that this formulation is not bounded and does not satisfy Property **III**. Therefore, a simple probabilistic formulation is not possible.

Although more elaborate conditions may enable a probabilistic formulation, it loses a degree of simple, intuitive appeal.

EXAMPLE 1. Let $\Omega=\{\omega_1,\ldots,\omega_{10}\}$, $A=\{\omega_1, \omega_2, \omega_3\}$, $B=\{\omega_2, \omega_3, \omega_4, \omega_5\}$; then $A\cap B=\{\omega_2,\omega_3\}$. If $P(\omega_i)=0.1$, $1\le i\le 10$, then $P(A)=0.3$, $P(B)=0.4$, and $P(A\cap B)=0.2$. Calculating $P(A|B)$ when A and B are fully independent $u(A,B)=1$, and fully dependent, $u(A,B)=0$, we get

$$P(A|B)=\begin{cases} P(A)=0.3, & \text{if } u(A,B)=1,\\ \dfrac{P(A\cap B)}{P(B)}=0.5, & \text{if } u(A,B)=0.\end{cases}$$

Let $C=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\}$, satisfying the condition that $B\subset C\subset\Omega$; then it should be true that $P(A\cap B)/P(B)>P(A\cap C)/P(C)>P(A)$, since $P(A\cap B)/P(B)>P(A)$. However, $P(A\cap C)/P(C)=0.6$, since $A\cap C=\{\omega_1,\omega_2,\omega_3\}$, $P(A\cap C)=0.3$, and $P(C)=0.5$. This means that $P(A\cap C)/P(C)$ is not between $P(A\cap B)/P(B)$ and $P(A)$. Property **III** is violated.

It can be further observed that when A and B are independent,

$$P(A)=\frac{P(A\cap\Omega)}{P(\Omega)}=\frac{P(A\cap B)+P(A\cap\bar{B})}{P(B)+P(\bar{B})}, \tag{2}$$

since $P(A\cap\Omega)=P(A\cap B)+P(A\cap\bar{B})$ and $P(\Omega)=P(B)+P(\bar{B})$. Using (1) and (2), we can redefine conditional probability as

$$P(A|B)=\frac{P(A\cap B)+u(A,B)P(A\cap\bar{B})}{P(B)+u(A,B)P(\bar{B})}.$$

When A and B are dependent $u(A,B)=0$, only the elements in B are relevant, limiting the probability of A. However, when A and B are independent, $u(A,B)=1$, B is not relevant, and does not limit the probability of A. Elements not in B are considered.

Let us now define $u'(A,B)$ as the *degree of independence* of events A and B. $u'(A,B)\in[0,1]$, $u'(A,B)=0$ when $u(A,B)=0$, $u'(A,B)=1$ when $u(A,B)=1$, and $u'(A,B)$ is monotonically nondecreasing as A and B become more independent. The definition is based on the *fuzzy* extension of binary independence to include the values between 0 and 1.

Extended conditional probability can now be defined as

$$P'(A|B) = \frac{P(A \cap B) + u'(A,B)P(A \cap \bar{B})}{P(B) + u'(A,B)P(\bar{B})} \tag{3}$$

by replacing $u(A,B)$ with $u'(A,B)$. Using $u'(A,B)$ to weight the probability of elements of Ω not in B is a new concept that avoids the problem of adding elements with full membership only. [If E is any event of which $\tilde{E}$ is a fuzzy subset, a fuzzy event [13], and the membership of all elements x of $\tilde{E}$ is uniform, $\mu_{\tilde{E}}(x) = c$, $\forall x$ where $c \in [0,1]$, then $P(\tilde{E}) = \Sigma_{x \in E}\, \mu_{\tilde{E}}(x)P(x) = \mu_{\tilde{E}}(x)\, \Sigma_{x \in E}\, P(x) = \mu_{\tilde{E}}(x)P(E)$.] That means that $u'(A,B)P(A \cap \bar{B})$ is equal to the probability of a fuzzy subset of $A \cap \bar{B}$, where $u'(A,B)$ is the uniform membership of all elements. As $u'(A,B)$ approaches 1, $P'(A|B)$ approaches $P(A)$, and as $u'(A,B)$ approaches 0, $P'(A|B)$ approaches $P(A \cap B)/P(B)$. The concept is illustrated in Figure 1.

The definition can be interpreted to mean that when $u'(A,B) = 0$, B stands out against its background $\bar{B}$, representing a significant factor with respect to A capable of altering $P'(A|B)$ when it changes. When $u'(A,B) = 1$, B blends into the background $\bar{B}$ becoming Ω, insignificant with respect to A, and incapable of effecting change. When $0 < u'(A,B) < 1$, B possesses a degree of significance with respect to A; the closer to 0, the more significant. When B changes, $P'(A|B)$ is altered to a degree; it is dampened by a partial dependence on B.

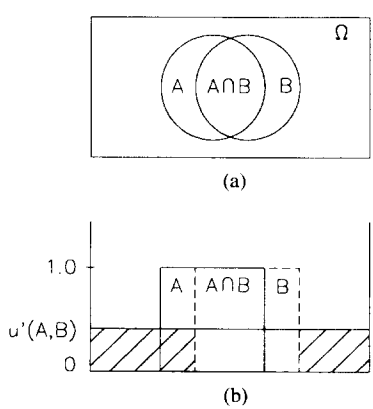

Fig. 1. (a) Probability space Ω containing events A and B. (b) Membership view of events A and B in probability space Ω. The hatched region shows the fuzzy subset of $\bar{B}$ with uniform degree of membership $u'(A,B)$.

Theorem 1 shows that the *fuzzy* formulation of ECP expressed by (3) satisfies the properties **I** through **IV-C** above, making it a valid formulation.

THEOREM 1. *Extended conditional probability as expressed in* (3) *satisfies properties* **I** *through* **IV-C**.

Proof. (1) Properties **I** and **II** are easily proven by substituting the appropriate value, 0 or 1, for $u'(A,B)$.

(2) When $P(A)<P(A\cap B)/P(B)$, then the ratio $P(A\cap\bar{B})/P(\bar{B})<P(A\cap B)/P(B)$. As $u'(A,B)$ increases from 0 to 1, the ratio contributes an increasingly larger amount to (3), causing $P'(A|B)$ to be monotonically nonincreasing. Property **IV-A** is satisfied. A similar argument proves Property **IV-B** where the ratio is reversed, and $P'(A|B)$ is monotonically nondecreasing. When $P(A)=P(A\cap B)/P(B)$, the ratio $P(A\cap\bar{B})/P(\bar{B})=P(A\cap B)/P(B)$. Since the ratio always contributes equally to the numerator and denominator, $P'(A|B)$ remains constant. Property **IV-C** is satisfied.

(3) Property **III** is implied by Properties **IV-A** and **IV-B**.

The preceding discussion and theorem pertained to the extension of conditional probability $P(A|B)$ based on the unconditional independence of A and B. Unconditional independence and conditional independence are not the same; one may exist without the other [6, 11]. However, the conditional probability $P(A|B,C)$ can likewise be extended. In this case, conditional independence is the independence of A and B given C. If we again use $u(A,B)$ to, in this case, express the conditional independence of A and B, conditional probability $P(A|B,C)$ can be expressed as

$$P(A|B,C)=\begin{cases} P(A|C), & \text{if } u(A,B)=1,\\ \dfrac{P(A\cap B\cap C)}{P(B\cap C)}, & \text{if } u(A,B)=0.\end{cases}$$

Similar to the previous observation concerning $P(A)$, it can be observed that

$$P(A|C)=\frac{P(A\cap C\cap\Omega)}{P(C\cap\Omega)}=\frac{P(A\cap C\cap(B\cup\bar{B}))}{P(C\cap(B\cup\bar{B}))}$$

$$=\frac{P(A\cap B\cap C)+P(A\cap\bar{B}\cap C)}{P(B\cap C)+P(\bar{B}\cap C)}.$$

This means that conditional independence can be redefined using $u(A, B)$:

$$P(A|B,C) = \frac{P(A \cap B \cap C) + u(A,B)P(A \cap \bar{B} \cap C)}{P(B \cap C) + u(A,B)P(\bar{B} \cap C)}.$$

Using degree of independence $u'(A, B)$ in place of binary independence, we can now define extended conditional probability under conditional independence:

$$P'(A|B,C) = \frac{P(A \cap B \cap C) + u'(A,B)P(A \cap \bar{B} \cap C)}{P(B \cap C) + u'(A,B)P(\bar{B} \cap C)}. \tag{4}$$

We see that (3) and (4) have the same form. As a result, (4), expressing extended conditional probability under conditional independence, satisfies Properties **I** through **IV-C**.

THEOREM 2. *Extended conditional probability as expressed in* (4) *satisfies properties* **I** *through* **IV-C**.

The proof of Theorem 2 is similar to that of Theorem 1.

5. ABSOLUTE EXTENDED CONDITIONAL PROBABILITY AND ITS APPROXIMATIONS

A special case of conditional probability $P(A|B)$ occurs when $B \subseteq A$. In this case, $A \cap B = B$, and when A and B are *not* independent, $P(A|B) = 1$. However, if A and B are independent, it still holds that $P(A|B) = P(A)$. Under these conditions, we define $P_A(A|B)$ as *absolute conditional probability*. Employing the standard definition of independence, absolute conditional probability assumes the values

$$P_A(A|B) = \begin{cases} 1, & \text{if } u(A,B) = 0, \\ P(A), & \text{if } u(A,B) = 1. \end{cases}$$

If, instead, we use degree of independence, then *absolute extended conditional probability* (AECP) under unconditional independence, as derived

from (3), is

$$P'_A(A|B) = \frac{P(B) + u'(A,B)P(A \cap \bar{B})}{P(B) + u'(A,B)P(\bar{B})}. \tag{5}$$

AECP requires knowing $P(B)$ and $P(A \cap \bar{B})$. More extensive probabilities are required when dealing with conditional independence. Therefore, approximations based on limited and accessible information, such as the a priori probability $P(A)$, would be helpful. Here we present two parameterized formulas to approximate the form of AECP, where m is a scaling parameter.

1. Exponential Degree of Independence (EDI) (based on a generalization of a formula presented by Hummel and Manevitz [5])

$$P'_A(A|B) = P(A)^{u'(A,B)^m} \tag{6}$$

for unconditional independence, and

$$P'_A(A|B,C) = P(A|C)^{u'(A,B)^m}$$

for conditional independence.

2. Linear Degree of Independence (LDI)

$$P'_A(A|B) = 1 - u'(A,B)^m[1 - P(A)] \tag{7}$$

for unconditional independence, and

$$P'_A(A|B,C) = 1 - u'(A,B)^m[1 - P(A|C)]$$

for conditional independence.

When $m = 0.5$, EDI fairly closely approximates AECP. When $m = 1$, LDI does not closely approximate AECP; however, it does have the simple intuitive appeal of a linear relationship. Figure 2 compares the AECP, (5), with the two approximations, (6) and (7), based on $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap \bar{B}) = 0.1$, and $P(\bar{B}) = 0.8$.

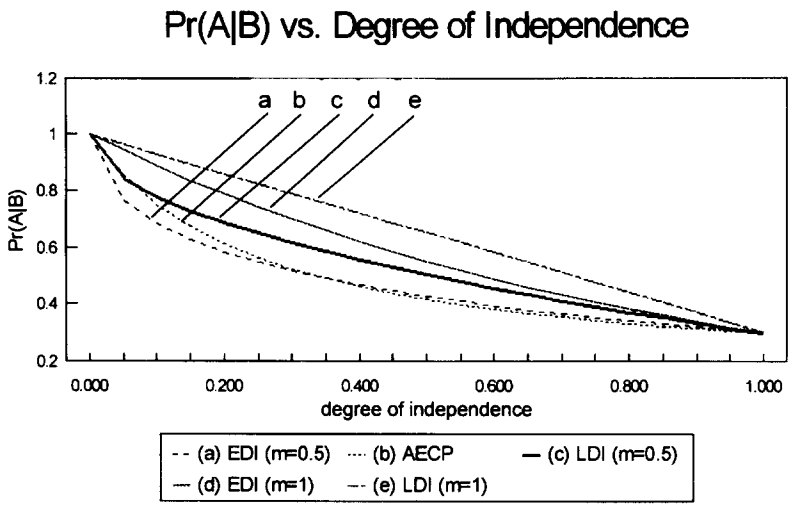

Fig. 2. A comparison of absolute extended conditional probability and approximations. $P(A)=0.3$, $P(B)=0.2$, $P(A\cap\overline{B})=0.1$, and $P(\overline{B})=0.8$.

6. APPLICATIONS OF EXTENDED CONDITIONAL PROBABILITY

6.1. LINGUISTIC INDEPENDENCE

Extended conditional probability takes a single point numeric value based on a precisely designated value for degree of independence. For example, the degree of independence of A and B may be determined to be 0.3, i.e., $u'(A,B)=0.3$. However, independence is not a precise concept; it has "a vague and intuitive sense" [8]. It is more reasonable to express degree of independence as *about 0.3* or *sort of independent*, capturing the vagueness and intuitivity. If, instead of a single point numeric value, degree of independence is a linguistic variable taking a linguistic value, such as **more_or_less_independent**, then degree of independence is defined by a fuzzy set over $[0,1]$. For example, **more_or_less_independent** may be defined as in Figure 3. Using degree of independence as a linguistic variable, ECP becomes a linguistic probability [14].

EXAMPLE 2. Let $P(A)=0.3$, $P(B)=0.2$, $P(A\cap\overline{B})=0.1$, and $P(\overline{B})=0.8$. Also, let degree of independence be a linguistic variable with the