Information Fusion 26 (2015) 136–143

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Jeffrey's rule of conditioning with various forms for uncertainty

Ronald R. Yager

Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 13 November 2014
Accepted 1 December 2014
Available online 9 December 2014

Keywords:
Conditional probability
Uncertainty
Measure
Jeffrey's rule
Possibility

ABSTRACT

We first introduce Jeffrey's rule of conditioning and explain how it allows us to determine the probability of an event related to one variable from information about a collection of conditional probabilities of that event conditioned on the state another variable. We note that in the original Jeffrey paradigm we have the uncertainty about the state of the conditioning variable expressed as a probability distribution. Here we extend this by allowing alternative formulations for the uncertainty about the conditioning variable. We first consider the case where our uncertainty is expressed in terms of a measure. This allows us to consider the case where our uncertainty is a possibility distribution. We next consider the case where our uncertainty about the conditioning variable is expressed in terms of a Dempster–Shafer belief structure. Finally we consider the case where we are ignorant about the underlying distribution and must use the decision maker's subjective attitude about the nature of uncertainty to provide the necessary information to use in the Jeffrey rule.

1. Introduction

The Jeffrey rule of conditioning was introduced by Jeffrey [1–4] and investigated by other researchers [5–13]. It provides a procedure for determining the probability of an event A associated with one variable, for example U, based upon a collection of conditional probabilities of A conditioned upon the state of another variable V. The conditioning states of V, B_j for $j = 1$ to q, form a partition of the space associated with V. For example A may correspond to presence of a particular disease and the variable V corresponds the results of a test and the conditionals are probability of having the disease for particular test result. Jeffrey's rule allows for some uncertainty about the knowledge of state of the variable V. Thus while we know the conditional probability of A given the state of V we are not sure of V. More specifically in the original Jeffrey work it is assumed that the knowledge about the state of V is carried by a probability distribution over the partitioning sets, the B_j. Here we provide an extension of the Jeffrey paradigm by considering that the information about the state of V is carried uncertainty formulations other then a probability distribution.

2. Jeffrey rule of conditional probabilities

Assume U is a variable taking its value in the space X and V is another variable taking its value in the space Y. Assume A is a subset of X whose probability is of interest to us. Let B_j, for $j = 1$ to q be a partition of the space Y, that is $B_j \cap B_k = \varnothing$ and $\bigcup_{j=1}^{q} B_j = Y$. Assume for $j = 1$ to q that $P(A|B_j)$ are the conditional probabilities that $U \in A$ given $V \in B_j$. If we know that $V \in B_j$ then the $P_U(A) = P(A|B_j)$. Jeffrey [1–4] provided a formula for determining the probability that $U \in A$, $P_U(A)$, in the case where we have some uncertainty with respect to the value of V. This formula has come to be known as Jeffrey's Rule of conditioning which we introduce in the following. If $P_V(B_j)$, for $j = 1$ to q, are a collection of probabilities so that $\sum_j P_V(B_j) = 1$ then Jeffrey suggested

$$P_U(A) = \sum_{j=1}^{q} P(A|B_j) P_V(B_j)$$

Jeffrey's original intention in introducing this rule was to provide a procedure for updating of our knowledge of the probability of A given the current information about the $P_V(B_j)$. A typical example of the use of this rule is illustrated in the following example.

Example. A manufacturer is deciding whether to expand his plant. His decision hinges on whether the expansion will be profitable. In the following we shall let A denote the event that the expansion will be profitable. From past experience he knows that the profitability of his expansion, A, depends on the future interest rates. He can partition the situation with respect to interest rates into three mutually exclusive sets, B_1 = interest rates decrease, B_2 = interest rates remain the same and B_3 = interest rates increase. Furthermore he knows from his experience that

E-mail address: yager@panix.com

http://dx.doi.org/10.1016/j.inffus.2014.12.001
1566-2535/

$P(A|B_1) = 0.8$
$P(A|B_2) = 0.5$
$P(A|B_3) = 0.3$

Unfortunately his knowledge about the future interest rates is uncertain. Based upon statements made by the members of the Federal Reserve Board, which are somewhat conflicting, he estimates the following probabilities

$P_V(B_1) = 0.4$
$P_V(B_2) = 0.4$
$P_V(B_3) = 0.2$

Using this information he applies Jeffrey's rule to determine the probability that the expansion will be profitable

$$P_U(A) = \sum_{j=1}^{3} P(A|B_j)P_V(B_j) = (0.8)(0.4) + (0.5)(0.4) + (0.3)(0.2) = 0.58$$

Thus in this example the probability the expansion will be profitable is 0.58.

An interesting property of the Jeffrey's rule is the following. Assume A_i, $i = 1$ to r is a partition of X, that is $A_i \cap A_j = \varnothing$ and $\bigcup_{i=1}^{r} A_i = X$. Furthermore in this case we have then for each A_i and B_j we have $P(A_i|B_j)$ where $\sum_{i=1}^{r} p(A_i|B_j) = 1$. Now from Jeffrey's rule we know for each A_i that

$$P_U(A_i) = \sum_{j=1}^{q} p(A_i|B_j)P_V(B_j)$$

From this we see that

$$\sum_{i=1}^{r} P_U(A_i) = \sum_{i=1}^{r}\left(\sum_{j=1}^{q} P(A_i|B_j)P_V(B_j)\right) = \sum_{j=1}^{q} P_V(B_j)\left(\sum_{i=1}^{r} P(A_i|B_j)\right)$$
$$= \sum_{j=1}^{q} P_V(B_j) = 1$$

Central to the use of the Jeffrey's rule is the assumption that the conditional probabilities, the $P(A|B_j)$, are fundamental objects of human perception of probabilistic uncertainty . We note that this is very much in the sprit of Bayesian modeling by in which conditional probability are seen as fundamental to the human organization of knowledge, a position clearly stated by Pearl in [14].

Here we shall view the Jeffrey rule more generally as a kind of weighted average, mean, of the conditional probabilities. Thus here we view

$$P_U(A) = \sum_{j=1}^{q} w_j P(A|B_j)$$

were the w_j are a set of weights having the properties that $w_j \in [0, 1]$ and $\sum_{j=1}^{q} w_j = 1$. In the case of the original Jeffrey rule, $w_j = P_V(B_j)$.

Given the properties of the weights we can observe some basic features of the mean formula just introduced [15]

(1) Boundedness

$\text{Min}_j[P(A|B_j] \leqslant P_U(A) \leqslant \text{Max}_j[P(A|B_j]$

(2) Monotonicity:
If $P(A|B_j]$ and $\widetilde{P}(A|B_j)$ are two sets of conditional probability so that

$P(A|B_j) \geqslant \widetilde{P}(A|B_j)$ for $j = 1$ to q

then $P_U(A) \geqslant \widetilde{P}_U(A)$.

(3) Idempotency
If all $P(A|B_j)$ are the same, all $P(A|B_j) = p$, then $P_U(A) = p$.

Once taking this perspective that we can view the determination of $P_U(A)$ as a mean of the conditional probabilities, the $P(A|B_j)$ weighted by the $P_V(B_j)$, we can look at other methods for this implementing mean.

One alternative is to use the median value. Here if we let h be an index function so that $h(k)$ is the index of the kth largest of the $P(A|B_j)$ then $P_U(A) = P(A|B_{h(k^*)})$ where k^* is such that

$$\sum_{k=1}^{k^*} P_V(B_{h(k)}) \geqslant 0.5 < \sum_{k=1}^{k^*-1} P_V(B_{h(k)}).$$

Here then we order the $P(A|B_j)$ in decreasing order of their associated $P_V(B_j)$ and then take $P_U(A)$ equal to the $P(A|B_j)$ where the ordered sum of the $P_V(B_j)$ crosses 0.5.

Let us apply this at our earlier example involving profitability and interest rates.

Example. Here we have $P(A|B_1) = 0.8$, $P(A|B_2) = 0.5$ and $P(A|B_j) = 0.3$ and $P(B_1) = 0.4$, $P_V(B_2) = 0.4$ and $P_V(B_3) = 0.2$. Here since the ordering the of $P(A|B_j)$ is

$P(A|B_1) > P(A|B_2) > P(A|B_3)$

we can use the following to calculate $P(A)$

$P(A\|B_j)$	$P_V(B_j)$	$\sum P_V(B_j)$
0.8	0.4	0.4
0.5	0.4	0.8 ⇐ Crossover
0.2	0.2	1.0

In this case we will have $P_U(A) = P(A|B_2) = 0.5$

3. Formulations of measure based uncertainty

In the preceding we have assumed that our knowledge about the uncertainty associated with the B_j events is available in the form of a probability distribution. In some cases our information about the uncertainty associated with B_j may be available in some other form. One very general framework for representing information about uncertainty is a measure. Here we shall consider an extension of the Jeffrey rule to the case of where our information about the uncertainty associated with the B_j is expressed in the form of a measure. We first review some ideas about measures and their role in modeling uncertain information [16,17].

A fuzzy measure μ on a space X is a mapping μ: $2^X \rightarrow [0, 1]$ such that

(1) $\mu(\varnothing) = \varnothing$
(2) $\mu(X) = 1$
(3) $\mu(A) \geqslant \mu(B)$ if $B \subseteq A$

The third property is called monotonicity. In the following we shall simply refer to μ as a measure.

Assume μ_1 and μ_2 are two measures on the space X such that $\mu_1(A) \geqslant \mu_2(A)$ for every A. We shall denote this as $\mu_1 \geqslant \mu_2$ and say that μ_1 is a more optimistic measure. We note that since $A \subseteq A \cup B$ and $B \subseteq A \cup B$ then $\mu(A \cup B) \geqslant \mu(A)$ and $\mu(A \cup B) \geqslant \mu(A)$ and hence $\mu(A \cup B) \geqslant \text{Max}[\mu(A), \mu(B)$. Similarly, since $A \cap B \subseteq A$ and $A \cap B \subseteq B$ then $\mu(A \cap B) \leqslant \mu(A)$ and $\mu(A \cap B) \leqslant \mu(B)$ and hence $\mu(A \cap B) \leqslant \text{Min}[\mu(A), \mu(B)]$.

We shall say a measure μ is superadditive if for all $A \cap B = \varnothing$ we have $\mu(A \cup B) \geqslant \mu(A) + \mu(B)$ and subadditive if $\mu(A \cup B) \leqslant \mu(A) + \mu(B)$