Membership Functions and Probability Measures of Fuzzy Sets

Nozer D. SINGPURWALLA and Jane M. BOOKER

The notion of fuzzy sets has proven useful in the context of control theory, pattern recognition, and medical diagnosis. However, it has also spawned the view that classical probability theory is unable to deal with uncertainties in natural language and machine learning, so that alternatives to probability are needed. One such alternative is what is known as "possibility theory." Such alternatives have come into being because past attempts at making fuzzy set theory and probability theory work in concert have been unsuccessful. The purpose of this article is to develop a line of argument that demonstrates that probability theory has a sufficiently rich structure for incorporating fuzzy sets within its framework. Thus probabilities of fuzzy events can be logically induced. The philosophical underpinnings that make this happen are a subjectivistic interpretation of probability, an introduction of Laplace's famous genie, and the mathematics of encoding expert testimony. The benefit of making probability theory work in concert with fuzzy set theory is an ability to deal with different kinds of uncertainties that may arise within the same problem.

KEY WORDS: Decision making; Expert testimony; Fuzzy control; Laplace's genie; Likelihood; Machine learning; Membership functions; Subjective probability.

1. PREAMBLE

Are probabilistic methods and statistical techniques the best available tools for solving problems involving uncertainty?

This question is often now being answered in the negative, especially by computer scientists and engineers. These respondents are motivated by the view that probability is inadequate for dealing with "certain kinds" of uncertainty. Thus alternatives are needed to fill the gap. One such alternative is Zadeh's (1978) *possibility theory*, the genesis of which lies in his theory of *fuzzy sets* (see Zadeh 1965). Because probability theory prescribes a calculus for the treatment of uncertainty about the outcomes of a random phenomenon, but not about the uncertainty of classification (or the placement of an outcome in a given class), Zadeh (1986b) has claimed that "probability lacks sufficient expressiveness to deal with uncertainty in natural language." In contrast, fuzzy set theory prescribes a calculus for the treatment of uncertainty associated with classification, or what has been termed "imprecision."

Because it is possible that both uncertainty and imprecision can be present in the same problem, Zadeh (1995) has also claimed that

> "probability must be used in concert with fuzzy logic to enhance its effectiveness. In this perspective, probability theory and fuzzy logic are complementary rather than competitive."

It is this statement that has motivated the thesis of this article.

Our aim here is to explore how probability theory and fuzzy set theory can be made to work in concert, so that uncertainty of outcomes and of imprecision can be treated in a unified and coherent manner. This we are able to do in several stages, the first one being a sharper appreciation of the key elements of probability theory; we do this in Section 2. In Section 3 we introduce the notion of a fuzzy set and a calculus for operations with fuzzy sets; this calculus is de facto a calculus for the treatment of imprecision. In Section 4 we review previous attempts to make probability theory and fuzzy set theory work in concert and discuss why these attempts have been unsatisfactory. Section 5 constitutes the crux of the article; here we describe our attempt at achieving the objective of Section 4, and present the philosophical and technical arguments that support our development. In Section 6 we illustrate how the material presented in Section 5 plays a role in the context of decision making in a fuzzy environment. In Section 7 we conclude the article by relating our work to some other interdisciplinary work on uncertainty quantification.

2. PROBABILITY: A CALCULUS FOR THE UNCERTAINTY OF OUTCOMES

The underlying set-theoretic premise for considering probability and its calculus is an experiment, $\mathcal{E}$, which is yet to be performed. Let x denote a generic outcome of $\mathcal{E}$, and let Ω denote the set of all conceived outcomes of $\mathcal{E}$; thus $x \in \Omega$. It is important to note that the probability theory does not tell one how to specify Ω; this choice is subjective and is up to the user. For convenience, we assume that Ω is a countable set. Let $\mathcal{F}$ denote a set whose members are subsets of Ω; that is, $\mathcal{F}$ is a family of sets. However, $\mathcal{F}$ is such that it contains Ω and ϕ, where ϕ is the null set. Furthermore, $\mathcal{F}$ is closed under unions and intersections; that is, if $A, B \in \mathcal{F}$, then $(A \cup B)$ and $(A \cap B)$. The subsets of Ω are called *events*, and in probability theory it is presumed that the events are well defined or "sharp" (also known as "crisp"); that is, there is no ambiguity in declaring whether any outcome x of Ω belongs to A or to its complement A^c. In contrast, with fuzzy sets there is ambiguity in classifying an x in a subset A or A^c, because A is not sharply defined; we discuss this in detail later. If the outcome of $\mathcal{E}$, say x, is such that $x \in A$, then we say that event A has occurred.

Nozer D. Singpurwalla is Professor of Statistics, The George Washington University, Washington, DC 20052 (E-mail: *nozer@gwu.edu*). Jane M. Booker is Technical Staff Member, Los Alamos National Laboratory, Los Alamos, NM 87545 (E-mail: *jmb@lanl.gov*). Discussions with Sallie Keller-McNulty, Dan Lunn, Mary Lunn, Timothy Ross, and Alyson Wilson resulted in this version of this manuscript. Sallie deserves a special mention for providing an environment under which this work was done. Dennis Lindley's comments (4 pages in all) caused the authors to rethink issues on the relationships between the rules of combining membership functions. The article by Laviolette et al. (1995) has facilited an appreciation of fuzzy methods. The authors applaud these authors for writing the article, and the editor of *Technometrics* for publishing it. Comments by two referees have enhanced its scope. Nozer D. Singpurwalla acknowledges the hospitality of The Santa Fe Institute, Santa Fe, New Mexico, during his tenure as a Visiting Professor in Fall 2001. His work was supported in part by U.S. Army Research Office grants DAAD19-01-1-0502 and DAAD19-02-1-0195. Support of the Engineering Sciences Division and the Statistical Sciences Group at LANL is also acknowledged. Last, but not least, the authors thank Russell Lenth, review editor of *JASA* for skillfully orchestrating the review and discussion of this article. This article is dedicated to Thomas Bement, whose vision initiated this work and who is no longer with us.

Journal of the American Statistical Association
September 2004, Vol. 99, No. 467, Review Article
DOI 10.1198/016214504000001196

Because $\mathcal{E}$ is yet to be performed, we are uncertain about the occurrence of any particular x. Consequently, we are also uncertain about the occurrence of event A. We describe this uncertainty by a number, $\mathcal{P}(A)$, where $0 \le \mathcal{P}(A) \le 1$; $\mathcal{P}(A)$ is the *probability of event* A, or the probability measure of the set A.

There are several interpretations of $\mathcal{P}(A)$; the one that is germane to our interest here is that $\mathcal{P}(A)$ is a two-sided bet (or wager) on the occurrence of event A. Specifically, $\mathcal{P}(A)$ is the amount that one is willing to stake out in exchange for a dollar should event A occur or, equivalently, $(1 - \mathcal{P}(A))$ is the amount staked in exchange for a dollar should event A not occur. Furthermore, the individual specifying $\mathcal{P}(A)$ is required to be indifferent between betting on A or A^c. The two-sided bet will be settled when $\mathcal{E}$ is performed and ω is observed, so that the disposition of A is known. An advantage of the foregoing interpretation of $\mathcal{P}(A)$ is that probability can be made "operational" via the mechanism of betting. This interpretation of probability is a basis for a personalistic (or a subjectivistic) theory of probability.

It is important to note that probability theory does not tell us how to arrive at a particular $\mathcal{P}(A)$, nor does it in its purely abstract form even attempt to interpret $\mathcal{P}(A)$. Many probabilists would declare that the assignment of initial probabilities is a job for a statistician, though some would say that the role of a statistician is to help clients formulate their prior knowledge, because it is the client who knows.

The calculus (or the algebra) of probability tells one how the various uncertainties (i.e., the initial probabilities) combine or *cohere*. In particular, if $\mathcal{P}(B)$ denotes the quantification of uncertainty of another event B, then

(a) $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$,

where

(b) $\mathcal{P}(A \cap B) = \begin{cases} 0 & \text{if } A \cap B = \phi \\ \mathcal{P}(A|B)\mathcal{P}(B) & \text{otherwise.} \end{cases}$

The quantity $\mathcal{P}(A|B)$ is called the *conditional probability* of A were B to occur. Like $\mathcal{P}(A)$, $\mathcal{P}(A|B)$ should lie between 0 and 1; it represents the amount that one is willing to stake on the event A should the event B occur but under the proviso that all bets on A will be called off should B not occur. It is crucial to bear in mind that $\mathcal{P}(A|B)$ is a bet in the subjunctive mood; this is because the disposition of B is unknown when $\mathcal{P}(A|B)$ is specified. Finally, ignoring the relevance of a conditioning event, events A and B are said to be *mutually independent* if $\mathcal{P}(A|B) = \mathcal{P}(A)$. The calculus given earlier has an axiomatic foundation based on behavioristic considerations.

Thus, to summarize, a foundation for the theory of probability is based on the following ingredients:

a. A well-defined set Ω and subsets of Ω.
b. An adherence to the "law of the excluded middle," the essential import of which is that any outcome ω of $\mathcal{E}$ belongs to a set A or to a set A^c, but not to both.
c. A calculus based on behavioristic axioms involving numbers between 0 and 1 that can be made operational once $\mathcal{E}$ is performed and its outcome observed.

3. FUZZY SET THEORY: A CALCULUS FOR IMPRECISION

The key point to note is that Zadeh (1965) introduced fuzzy set theory as a mathematical construct in set theory with no intention of using it to enhance, complement, or replace probability theory. The sometimes-held perception that fuzzy set theory is a substitute for probability theory is not correct! Indeed, Zadeh's article titled "Probability Measures of Fuzzy Events" (Zadeh 1968) suggested that, his intent was to simply expand the scope of applicability of probability theory to include fuzzy sets. But what are fuzzy sets and do we need them? In the next section we attempt to address these questions.

3.1 Fuzzy Sets and Membership Functions

To appreciate the nature of a fuzzy set, let us consider the following hypothetical example taken from Laviolette, Seaman, Barrett, and Woodall (1995). Let $\mathcal{X}$ denote the set of integers between 0 and 10, both inclusive; that is,

$$\mathcal{X} = \{0, 1, \ldots, 10\}.$$

Suppose that we are interested in a subset $\widetilde{A}$ of $\mathcal{X}$, where $\widetilde{A}$ contains all of the "medium" integers of $\mathcal{X}$. Thus

$$\widetilde{A} = \{x; x \in \mathcal{X} \text{ and } x \text{ is "medium"}\}.$$

For the purposes of this section, x is not to be viewed as an outcome of any experiment. Some reasons as to why one would be interested in sets like $\widetilde{A}$ are given in Section 3.4. Clearly, to be able to specify $\widetilde{A}$, we must be precise as to what we mean by a medium integer; that is, we must be able to operationalize (Deming 1986, p. 276) the term "medium integer." Whereas most would agree that 5 is a medium integer, what is the disposition of an integer like 7? Is 7 a medium integer, or is it a large integer? Our uncertainty (or vagueness) about classifying 7 as a member of the subset $\widetilde{A}$ makes $\widetilde{A}$ a fuzzy set. The uncertainty of classification arises because the boundaries of $\widetilde{A}$ are not sharp. The subset $\widetilde{A}$ rejects the law of the excluded middle, because an integer like 7 can simultaneously belong to and not belong to $\widetilde{A}$.

Membership functions were introduced as a way of dealing with the foregoing form of uncertainty of classification. Specifically, the number $m_{\widetilde{A}}(x)$ which lies between 0 and 1, reflects an assessor's view of the extent to which $x \in \widetilde{A}$. As a function of x, $m_{\widetilde{A}}(x)$ is known as the *membership function* of set $\widetilde{A}$. Clearly, the membership function is subjective, because it is specific to an individual assessor or a group of assessors. We also assume that for each $x \in \mathcal{X}$, the assessor is able to assign an $m_{\widetilde{A}}(x)$, and that this can be done for all subsets of the type $\widetilde{A}$ that are of interest.

If $m_{\widetilde{A}}(x) = 1$ (or 0) for all $x \in \mathcal{X}$, then $\widetilde{A}$ is the usual well-defined sharp (or crisp) set. Thus the notion of fuzzy sets incorporates that of crisp sets as a special case, and because it is on crisp sets that probability measures have been defined, our aim here is to develop a foundation for constructing probability measures of fuzzy sets.

3.2 The Calculus (or Algebra) of Fuzzy Sets

The membership function provides us with a vehicle for developing operations with fuzzy sets, such as unions, denoted by "∪", and intersections, denoted by "∩". Specifically, let $\widetilde{A}$ and $\widetilde{B}$ be two fuzzy sets with respective membership functions $m_{\widetilde{A}}(x)$ and $m_{\widetilde{B}}(x)$. Then (Zadeh 1965), $\forall x$,

$$m_{\widetilde{A}\cup\widetilde{B}}(x) \stackrel{\text{def}}{=} \max\left[m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)\right],$$

$$m_{\widetilde{A}\cap\widetilde{B}}(x) \stackrel{\text{def}}{=} \min\left[m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)\right],$$

and

$$m_{\widetilde{A}^c}(x) \stackrel{\text{def}}{=} 1 - m_{\widetilde{A}}(x).$$

Thus $\widetilde{A} \cup \widetilde{B}$ is that fuzzy set whose membership function is $m_{\widetilde{A}\cup\widetilde{B}}(x)$. The fuzzy sets $\widetilde{A} \cap \widetilde{B}$ and $\widetilde{A}^c$ are defined similarly.

Furthermore, if, $\forall x$,

$$m_{\widetilde{A}}(x) = m_{\widetilde{B}}(x), \quad \text{then } \widetilde{A} = \widetilde{B} \quad \text{(and vice versa)},$$

where "=" denotes the equality, and if, $\forall x$,

$$m_{\widetilde{A}}(x) \le m_{\widetilde{B}}(x), \quad \text{then } \widetilde{A} \subseteq \widetilde{B} \quad \text{(and vice versa)},$$

where "⊂" denotes a subset. It is of interest to note that fuzzy sets are defined only in terms of membership functions, whereas sharp sets are often discussed without any mention of indicator functions.

Zadeh (1965) also introduced the notion of the product and the sum of fuzzy sets $\widetilde{A}$ and $\widetilde{B}$, denoted by $\widetilde{A} \bullet \widetilde{B}$ (not to be confused with the intersection $\widetilde{A} \cap \widetilde{B}$) and $\widetilde{A} + \widetilde{B}$ as fuzzy sets whose membership functions are given as

$$m_{\widetilde{A}\bullet\widetilde{B}}(x) = m_{\widetilde{A}}(x) \cdot m_{\widetilde{B}}(x)$$

and

$$m_{\widetilde{A}+\widetilde{B}}(x) = m_{\widetilde{A}}(x) + m_{\widetilde{B}}(x) - m_{\widetilde{A}}(x) \cdot m_{\widetilde{B}}(x) \qquad \forall x.$$

Bellman and Giertz (1973) have claimed that of the foregoing relationships, $m_{\widetilde{A}\cup\widetilde{B}}(x)$ and $m_{\widetilde{A}\cap\widetilde{B}}(x)$ are not only natural, but also the only ones possible. They also claim that $m_{\widetilde{A}^c}(x)$ "appears to be reasonable," though not unique. We agree with this assessment but via arguments that are probabilistic; see Section 5.5.

Although the sum and product rules are intriguing, they are not standard from the standpoint of set theory; we are unable to interpret these rules. They could have been introduced by Zadeh in his quest for developing probability-like measures for fuzzy sets. We discuss this further, in Section 4.

3.3 Interpreting the Membership Function

The first point to note is that, like $\mathcal{P}(A)$, the probability of a set A, fuzzy set theory does not tell us how to specify $m_{\widetilde{A}}(x)$, the membership function of a fuzzy set $\widetilde{A}$. The second point to note is that whereas there is a logical requirement that $\mathcal{P}(A) \in [0, 1]$, the fact that $m_{\widetilde{A}}(x) \in [0, 1]$ is simply a convenience of scaling. The third point to note is that whereas $\mathcal{P}(A)$ can be interpreted as a two-sided bet (which in principle can be settled when A reveals itself), $m_{\widetilde{A}}(x)$ reflects an individual's view of the extent to which $x \in \widetilde{A}$; thus $m_{\widetilde{A}}(x)$ cannot be made operational in the same sense as $\mathcal{P}(A)$. Finally, it is *not* a requirement that $\sum_x m_{\widetilde{A}}(x)$ be 1, and thus $m_{\widetilde{A}}(x)$ as a function of x cannot be interpreted as a probability or, for that matter, as a conditional probability, as was done by Loginov (1966) and also by Barrett and Woodall (1997). How then can we interpret the membership function $m_{\widetilde{A}}(x)$?

Because $m_{\widetilde{A}}(x)$, as a function of x, reflects the extent to which $x \in \widetilde{A}$ [i.e., $m_{\widetilde{A}}(x)$ is an indicator of how likely it is that $x \in \widetilde{A}$], we may view $m_{\widetilde{A}}(x)$ as the *likelihood* of x for a fixed (i.e., specified) $\widetilde{A}$. Recall that even though the interpretation of a likelihood is almost always derived from a probability model, the likelihood is not a probability (in particular, it does not obey the addition rule) and in statistical inference, the likelihood function reflects the relative degrees of support that a fixed observation provides to several hypotheses. Furthermore, the specification of a likelihood is subjective. Thus our interpretation of the membership function is that it is a likelihood function with $\widetilde{A}$ taking the role of a fixed observation and the values of x taking the role of the hypotheses. To statisticians specializing in inference, our interpretation of the membership function as a likelihood will appear to be unconventional. This is because in the context of inference, the likelihood entails a fixed observation and a varying parameter. However, our structure for the likelihood is a consequence of the notion of the likelihood from a more philosophical viewpoint, and what we have proposed is in keeping with the foundational notion of a likelihood (see Basu 1975).

The foregoing points are best illustrated via the following example involving two fuzzy sets $\widetilde{A}$ and $\widetilde{B}$, where

$$\widetilde{A} = \{x : x \in \mathcal{X} \text{ and } x \text{ is "medium"}\}$$

and

$$\widetilde{B} = \{x : x \in \mathcal{X} \text{ and } x \text{ is "small"}\};$$

as before, $\mathcal{X} = \{0, 1, \ldots, 10\}$.

Suppose that an assessor assigns the membership functions $m_{\widetilde{A}}(x)$ and $m_{\widetilde{B}}(x)$ given in Table 1.

Clearly, $\sum_x m_{\widetilde{A}}(x)$ and $\sum_x m_{\widetilde{B}}(x)$ are not 1, nor is it true that $m_{\widetilde{A}}(x) + m_{\widetilde{B}}(x)$ is necessarily 1. A plot of $m_{\widetilde{A}}(x)$ and $m_{\widetilde{B}}(x)$ as a function of x, with $m_{\widetilde{A}}(x)$ and $m_{\widetilde{B}}(x)$ viewed as likelihoods is shown in Figure 1. The plots reflect the extent to which an x belongs to the sets $\widetilde{A}$ and $\widetilde{B}$.

Table 1. Membership Functions of $\widetilde{A}$ and $\widetilde{B}$

x	$m_{\widetilde{A}}(x)$	$m_{\widetilde{B}}(x)$	$m_{\widetilde{A}}(x) + m_{\widetilde{B}}(x)$
0	0	1	1
1	0	1	1
2	.2	.8	1
3	.5	.5	1
4	.8	.3	1.1
5	1	.1	1.1
6	.8	0	.8
7	.5	0	.5
8	.2	0	.2
9	0	0	0
10	0	0	0
Col. sums	4	3.7	–

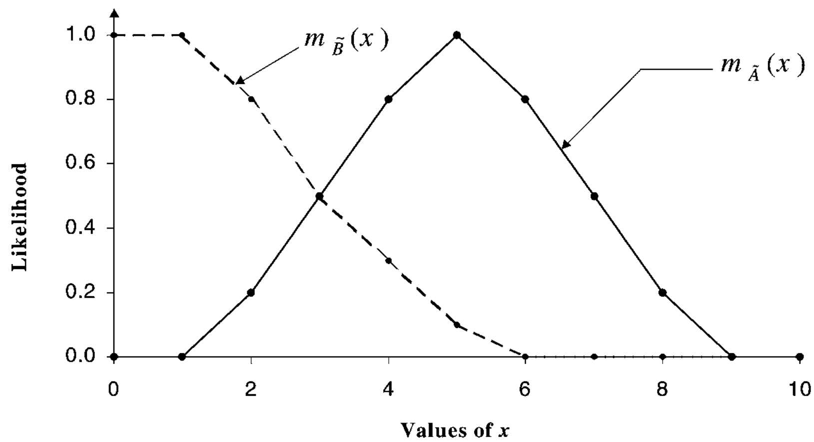

Figure 1. Membership Functions of $\widetilde{A}$ and $\widetilde{B}$.

3.4 Interest in Fuzzy Sets

Fuzzy sets were introduced by Zadeh (1965) based on the premise that an exact description of many real world situations is virtually impossible, and that imprecisely defined "classes" play an important role in human thinking and natural language. Examples are commonly used adjectives, such as "substantial," "significant," "accurate," "approximate," "small," and "medium." Furthermore, such activities as the communication of information, speech recognition, knowledge representation, medical diagnosis, and assessment of rare events suggest that the human brain often reasons with vague assertions, a fact that one needs to accept and to adjust to. All the same, computers that permeate our everyday lives do not reason as brains do, and it has been argued that the main distinction between human intelligence and machine intelligence lies in the ability of humans to manipulate imprecise concepts and imprecise instructions. The notion of fuzzy sets strives to balance exactness and simplicity in such a way that complexity can be reduced without oversimplification.

Imprecision can also arise in the context of real world decision making wherein the goals, the constraints, and the consequences of actions cannot be precisely specified. Whereas the human mind is able to cope with such impressions, normative decision theory (which provides a foundation for classical control theory) presumes a precise delineation of actions, outcomes, and consequences. Fuzzy sets provide a trade-off between the reality and the requirement, mentioned earlier, and in the context of decision making under uncertainty manifests under the label "fuzzy control." Indeed, it is in the context of control theory that the notion of fuzzy sets has had its biggest impact. For example, according to some estimates, in 1992 Japan produced about \$2 billion worth of products with control mechanisms considered fuzzy (see Kosko and Isaka 1993). Section 6 gives a flavor of fuzzy control, albeit in a narrow context.

4. PROBABILITY AND FUZZY SET THEORY IN CONCERT: PREVIOUS WORK

The earliest attempt at making probability and fuzzy set theory work in concert was made by Loginov (1966), who interpreted the membership function as a frequentist conditional probability. Specifically, if an experiment $\mathcal{E}$ were to yield an outcome x, then $m_{\widetilde{A}}(x)$ would be the probability that x is classified as a member of $\widetilde{A}$, that is,

$$m_{\widetilde{A}}(x) = \mathcal{P}(x \in \widetilde{A} | \mathcal{E} \text{ yields } x).$$

For a frequentist interpretation of this probability, Loginov conceptualized an infinite-sized *ensemble* of membership function specifiers, each of whom has to vote on $x \in \widetilde{A}$ or $x \notin \widetilde{A}$. Barrett and Woodall (1997), in introducing a probabilistic alternative to fuzzy logic controllers, appeared to follow a similar line of thought.

Zadeh (1995) dismissed Loginov's interpretation on grounds that requiring each voter to classify an observed x in $\widetilde{A}$ or $\widetilde{A}^c$ is unnatural, because fuzzy sets reject the law of the excluded middle. We concur with this criticism provided by Zadeh. Also, because membership functions are often specified by one individual based on subjective considerations, the consensus model involving an ensemble of voters is untenable.

The second attempt at making probability theory and fuzzy set theory work in concert was made by Zadeh (1968), in his imaginative but flawed article titled "Probability Measures of Fuzzy Events." His construction proceeds along the following lines. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a "probability measure space" with Ω, $\mathcal{F}$, and $\mathcal{P}$ as defined in Section 2. Recall that x, an outcome of $\mathcal{E}$, is a member of Ω, and assume for now that Ω is a countable set. Consider now a crisp set A, where $A \in \mathcal{F}$, and let $I_A(x)$ be the indicator of A, that is, $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Then it is easy to see that

$$\mathcal{P}(A) = \sum_x I_A(x)\mathcal{P}(x), \qquad x \in \Omega,$$

where $\mathcal{P}(x)$ is the probability that the outcome of $\mathcal{E}$ is x. An analog of the foregoing result when Ω is not countable is a relationship of the form

$$\mathcal{P}(A) = \int_\Omega I_A(x)\, d\mathcal{P}(x).$$

Motivated by this (well-known) result, plus the fact that $I_A(x)$ is itself a membership function, Zadeh has *declared* that the probability measure of a fuzzy subset $\widetilde{A}$ of Ω, which he calls a *fuzzy event*, is

$$\Pi(\widetilde{A}) = \int_\Omega m_{\widetilde{A}}(x)\, d\mathcal{P}(x) = \mathbf{E}\big[m_{\widetilde{A}}(X)\big], \tag{1}$$

where $m_{\widetilde{A}}(x)$ is the membership function of $\widetilde{A}$ and $\mathbf{E}$ denotes expectation. The point to be emphasized here is that the expectation is taken with respect to the initial probability measure $\mathcal{P}$ that has been defined on the (crisp) sets of Ω.

Having defined $\Pi(\widetilde{A})$ as before, Zadeh proceeded to show that

(a) $\widetilde{A} \subseteq \widetilde{B} \Rightarrow \Pi(\widetilde{A}) \leq \Pi(\widetilde{B})$,

(b) $\Pi(\widetilde{A} \cup \widetilde{B}) = \Pi(\widetilde{A}) + \Pi(\widetilde{B}) - \Pi(\widetilde{A} \cap \widetilde{B})$,

and

(c) $\Pi(\widetilde{A} + \widetilde{B}) = \Pi(\widetilde{A}) + \Pi(\widetilde{B}) - \Pi(\widetilde{A} \bullet \widetilde{B})$,

where we recall that $\widetilde{A} \bullet \widetilde{B}$ is the product (not the intersection) of $\widetilde{A}$ and $\widetilde{B}$. Extensions of the foregoing for the cases of finite and countable additivity follow. Finally, $\widetilde{A}$ and $\widetilde{B}$ are declared to be independent if

$$\Pi(\widetilde{A} \bullet \widetilde{B}) = \Pi(\widetilde{A}) \cdot \Pi(\widetilde{B}),$$

and the conditional probability of $\widetilde{A}$, were $\widetilde{B}$ to occur, denoted by $\Pi(\widetilde{A}|\widetilde{B})$, is defined as

$$\Pi(\widetilde{A}|\widetilde{B}) = \frac{\Pi(\widetilde{A} \bullet \widetilde{B})}{\Pi(\widetilde{B})} \quad \text{if} \quad \Pi(\widetilde{B}) \geq 0.$$

Thus when $\widetilde{A}$ and $\widetilde{B}$ are independent,

$$\Pi(\widetilde{A}|\widetilde{B}) = \Pi(\widetilde{A}).$$

Whereas the definition (1) has the virtue that when $\widetilde{A}$ is a crisp set, $\Pi(\widetilde{A}) = \mathcal{P}(\widetilde{A})$, so that the measure Π can be seen as a generalization of the measure $\mathcal{P}$, the question still remains as to whether Π is a probability measure. Properties (a), (b), and (c) seem to suggest that Π could indeed be viewed as a probability measure, but such a conclusion would be premature. This is because of the following arguments:

a. With property (b), the evaluation of $\Pi(\widetilde{A})$ and $\Pi(\widetilde{B})$ is enough to evaluate $\Pi(\widetilde{A} \cap \widetilde{B})$, whereas with probability, the evaluation of $\mathcal{P}(A)$ and $\mathcal{P}(B)$ is not sufficient to evaluate $\mathcal{P}(A \cap B)$, unless A and B are independent. Note that because

$$\Pi(\widetilde{A} \cap \widetilde{B}) = \mathbf{E}\big[m_{\widetilde{A}\cap\widetilde{B}}(X)\big] = \mathbf{E}\big[\min\big(m_{\widetilde{A}}(X), m_{\widetilde{B}}(X)\big)\big],$$

it can be easily seen that for $\mathcal{A} = \{x : m_{\widetilde{A}}(x) \leq m_{\widetilde{B}}(x)\}$

$$\Pi(\widetilde{A} \cup \widetilde{B}) = \int_{x\in\mathcal{A}} m_{\widetilde{B}}(x)\, d\mathcal{P}(x) + \int_{x\in\mathcal{A}^c} m_{\widetilde{A}}(x)\, d\mathcal{P}(x).$$

b. Property (c) has no analog in probability, because the notions of $(\widetilde{A} + \widetilde{B})$ and $(\widetilde{A} \bullet \widetilde{B})$ are not part of classical set theory. More important, conditional probability has only been defined in terms of $(\widetilde{A} \bullet \widetilde{B})$.

Given that the genesis of (1) is the probability measure space $(\Omega, \mathcal{F}, \mathcal{P})$, there is no law of probability that leads to this equation. Recall that because $m_{\widetilde{A}}(x)$ is not a probability, we may not appeal to the law of total probability to claim that the probability of a fuzzy set is

$$\int_{\Omega} m_{\widetilde{A}}(x)\, d\mathcal{P}(x).$$

Because of the foregoing concerns, it is our view that Zadeh's (1968) attempt at making fuzzy set theory and probability theory work in concert has also been unsuccessful. In what follows we propose a line of argument that is able to achieve Zadeh's goal.

5. THE SIMULTANEOUS TREATMENT OF IMPRECISION AND UNCERTAINTY: A NORMATIVE APPROACH

For the simultaneous treatment of imprecision and uncertainty, we conceptualize a scenario involving a wholeheartedly subjectivistic analyst, say $\mathcal{D}$ (in honor of de Finetti), who is interested in assessing the probability of a fuzzy set $\widetilde{A}$. That is, $\mathcal{D}$ needs to specify

$$\mathcal{P}_{\mathcal{D}}(\widetilde{A}) = \mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}),$$

where the generic X denotes the uncertain outcome of an experiment $\mathcal{E}$ and the subscript $\mathcal{D}$ denotes the fact that what is being assessed is $\mathcal{D}$'s personal probability, that is $\mathcal{D}$'s willingness to bet, as described in Section 2. We assume for now that $\mathcal{D}$ has no access to any membership function of $\widetilde{A}$ or is unwilling to consider an $m_{\widetilde{A}}(x)$, should one be available. The matter of $\mathcal{D}$ incorporating a membership function in his or her analysis is considered in Section 5.3.

With the fuzzy set $\widetilde{A}$ entering the picture, $\mathcal{D}$ is confronted with *two* uncertainties, one about the outcome $X = x$, and the other about the membership of x in $\widetilde{A}$. If $\widetilde{A}$ is a crisp set (as is normally the case in standard probability theory), then $\mathcal{D}$ would be confronted with only one uncertainty, namely the uncertainty that $X = x$.

As a subjectivist, $\mathcal{D}$ views imprecision as simply another uncertainty, and to $\mathcal{D}$ all uncertainties can be quantified only by probability. Thus $\mathcal{D}$ specifies two probabilities:

(a) $\mathcal{P}_{\mathcal{D}}(x)$, which is $\mathcal{D}$'s prior probability that an outcome of $\mathcal{E}$ will be x,

and

(b) $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, which is $\mathcal{D}$'s prior probability that an outcome x belongs to $\widetilde{A}$.

Whereas the specification of $\mathcal{P}_{\mathcal{D}}(x)$ is a standard operation, the assessment of $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ raises an issue. Specifically, because $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ is $\mathcal{D}$'s personal probability that x is classified in $\widetilde{A}$, the question arises as to who is doing the classification and on what basis such classification is done. Recall that Zadeh's criticism of Loginov's approach has been an individual's inability to classify an outcome in a fuzzy set.

The viewpoint that we adopt here is a philosophical one, as described later. This view has a historical precedence in Laplace's interpretation of probability (see Gigerenzer et al. 1990, p.11). An appreciation of this viewpoint is crucial to the development that we propose.

5.1 Laplace's Genie for Classification in Fuzzy Sets

Central to our development is the notion that *nature* is able to classify any x in any set $\widetilde{A}$ or $\widetilde{A}^c$, with precision, so that to nature there is no such a thing as a fuzzy set; all sets are crisp. That is, to nature the membership function for any fuzzy set $\widetilde{A}$

is of the form $m_{\widetilde{A}}(x) = 1$ or 0. Thus fuzzy sets are only a manifestation of *our* uncertainty about the boundaries of sharp sets. Consequently, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ is merely a reflection of $\mathcal{D}$'s uncertainty (or partial knowledge) of the boundaries of a crisp set. Nature will never reveal these boundaries, so $\mathcal{D}$'s uncertainty of classification is impossible to ever resolve.

The foregoing idealization is based on Laplace's famous *genie* (see Gigerenzer et al. 1990, p. 11), who knows it all and rarely tells it all. The genie is able to classify x with precision, but $\mathcal{D}$ is unsure of this classification. All the same, $\mathcal{D}$ has partial knowledge of the genie's actions, and this is encapsulated in $\mathcal{D}$'s $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$.

Laplace invoked the genie in his interpretation of probability. Like Newton, who preceded him, Laplace was a "determinist." Thus to Laplace, probability was merely a reflection of a human's partial knowledge about the behavior of nature that is fully deterministic. To Laplace's genie, there is no such thing as probability.

Irrespective of how one chooses to interpret $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, an assessment of this quantity is essential for developing a normative approach for assessing probability measures of fuzzy sets. In introducing $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, we have de facto reaffirmed Lindley's (1987) claim that probability is able to handle any situation that fuzzy logic can.

5.2 Making Classification Probabilities Operational

Before proceeding further with regard to $\mathcal{D}$'s assessment of $\mathcal{P}_{\mathcal{D}}(\widetilde{A})$, we need to resolve the following issue, which is germane vis-à-vis $\mathcal{D}$'s position as a subjectivist and $\mathcal{D}$'s inability to settle a bet based on $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$. Recall that nature does not ever reveal the classification (or not) of any x in an $\widetilde{A}$. Does the inability to settle a bet about classification make a subjectivistic specification of $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ a vacuous exercise?

Our answer to the foregoing question is "no." This is because one assesses probabilities to embed them within the general framework of decision making under uncertainty, and bets can be settled based on the decisions that one makes. Thus, as long as probability measures of fuzzy sets are harnessed within the broader framework of decision making under uncertainty, the exercise of assessing them is a useful one. Fortunately, this viewpoint is also supported in practice, because the biggest impact of fuzzy set theory has been in "fuzzy control," de facto decision making in an uncertain environment characterized by a state (and action) space that is a fuzzy set.

With the foregoing arguments in place, once $\mathcal{P}_{\mathcal{D}}(x)$ and $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ have been specified by $\mathcal{D}$ for all $x \in \Omega$, $\mathcal{D}$ will use the law of total probability to write

$$\begin{aligned}\mathcal{P}_{\mathcal{D}}(\widetilde{A}) &= \mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}) \\ &= \sum_x \mathcal{P}_{\mathcal{D}}(X \in \widetilde{A} | X = x)\mathcal{P}_{\mathcal{D}}(x) \\ &= \sum_x \mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})\mathcal{P}_{\mathcal{D}}(x),\end{aligned} \tag{2}$$

which is the expected value of $\mathcal{D}$'s classification probability with respect to $\mathcal{D}$'s prior probability of X.

We thus have a probability measure for a fuzzy set $\widetilde{A}$ that can be justified on the basis of personal (i.e., subjective) probabilities and the notion that probability is a reflection of one's partial knowledge about an event of interest. Equation (2) is based on $\mathcal{D}$'s inputs alone. The membership function $m_{\widetilde{A}}(x)$, which is the mainstay of fuzzy set theory, has yet to play a role. This role is articulated in Section 5.3.

5.3 Membership Functions as Expert Testimonies

To incorporate the role of membership functions in the assessment of a probability measure of a fuzzy set $\widetilde{A}$, we find it convenient to introduce a new character into our analysis–namely an expert, say $\mathcal{Z}$ (in honor of Zadeh), whose expertise lies in specifying a membership function $m_{\widetilde{A}}(x)$ for all $x \in \Omega$, and a fuzzy set $\widetilde{A}$. Suppose, then, that after assessing $\mathcal{P}_{\mathcal{D}}(\widetilde{A})$, via (2), $\mathcal{D}$ consults $\mathcal{Z}$ and elicits from $\mathcal{Z}$, the function $m_{\widetilde{A}}(x)$ $\forall x \in \Omega$. $\mathcal{D}$ now needs to update his or her $\mathcal{P}_{\mathcal{D}}(\widetilde{A})$ in light of $\mathcal{Z}$'s expert testimony $m_{\widetilde{A}}(x)$, but in a manner consistent (i.e., coherent) with $\mathcal{P}_{\mathcal{D}}(\widetilde{A})$, which encapsulates $\mathcal{D}$'s prior opinions about the probability of $\widetilde{A}$. $\mathcal{D}$ will therefore appeal to the calculus of probability, and, following standard practice, consider (via the law of total probability) the proposition

$$\begin{aligned}&\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A} | m_{\widetilde{A}}(x)\big) \\ &\quad = \sum_x \mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A} | m_{\widetilde{A}}(x), X = x\big)\mathcal{P}_{\mathcal{D}}\big(X = x | m_{\widetilde{A}}(x)\big).\end{aligned} \tag{3}$$

In writing out the foregoing, $\mathcal{D}$ treats $m_{\widetilde{A}}(x)$ as an unknown quantity, the supposition being that when $\mathcal{D}$ is contemplating his or her prior probabilities, $\mathcal{Z}$'s response is unknown. If $\mathcal{D}$ assumes (and reasonably so) that the process by which X is generated is independent of the manner by which $\mathcal{Z}$ specifies $m_{\widetilde{A}}(x)$; then (3) becomes

$$\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A} | m_{\widetilde{A}}(x)\big) = \sum_x \mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A} | m_{\widetilde{A}}(x)\big)\mathcal{P}_{\mathcal{D}}(x). \tag{4}$$

In writing out the foregoing, we have assumed that in $m_{\widetilde{A}}(x)$, only the value at x affects $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} | m_{\widetilde{A}}(x))$.

To proceed further, $\mathcal{D}$ must evaluate $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} | m_{\widetilde{A}}(x))$, and for this $\mathcal{D}$ appeals to Bayes's law; specifically,

$$\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A} | m_{\widetilde{A}}(x)\big) \propto \mathcal{P}_{\mathcal{D}}\big(m_{\widetilde{A}}(x) | x \in \widetilde{A}\big)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}). \tag{5}$$

But in actuality, $m_{\widetilde{A}}(x)$ is known to $\mathcal{D}$, as $\mathcal{Z}$'s expert testimony. Thus the middle term becomes a likelihood, and (5) gets written as

$$\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \propto \mathcal{L}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}); \tag{6}$$

here $\mathcal{L}_{\mathcal{D}}(x \in \widetilde{A};\ m_{\widetilde{A}}(x))$ is $\mathcal{D}$'s likelihood that $\mathcal{Z}$ specifies $m_{\widetilde{A}}(x)$, were nature to classify x as belonging to $\widetilde{A}$. In writing the foregoing, we have used the convention that $\mathcal{D}$ knows all terms on the right side of the semicolon at the time of making the relevant assessments.

The scenario described here parallels that encountered in the literature on the probabilistic treatment of expert testimony (see, e.g., Lindley 1983). It is helpful to remark that our development entails two likelihoods. The first is $m_{\widetilde{A}}(x)$, which is $\mathcal{Z}$'s likelihood, revealed to $\mathcal{D}$ as expert testimony; it is personal to $\mathcal{Z}$ in the sense that it is subjectively specified by $\mathcal{Z}$. The second is $\mathcal{L}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x))$, which is personal to $\mathcal{D}$ and is based on $\mathcal{D}$'s assessment of the expertise of $\mathcal{Z}$ on how nature would classify an x in $\widetilde{A}$. It encapsulates $\mathcal{D}$'s opinion of what $\mathcal{Z}$ would say were $x \in \widetilde{A}$.

If $\mathcal{D}$ wishes to adopt $\mathcal{Z}$'s input without any modification, then $\mathcal{D}$ would set, $\forall x \in \Omega$,

$$\mathcal{L}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) = m_{\widetilde{A}}(x), \tag{7}$$

and then (4), rewritten to account for the fact that $m_{\widetilde{A}}(x)$ is known to $\mathcal{D}$, becomes

$$\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \propto \sum_x m_{\widetilde{A}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})\mathcal{P}_{\mathcal{D}}(x); \tag{8}$$

this is $\mathcal{D}$'s nonnormalized probability measure of the fuzzy set $\widetilde{A}$.

If $\mathcal{D}$'s judgment is such that $\mathcal{Z}$ tends to exaggerate (or to underestimate) the membership of a particular x in $\widetilde{A}$, then $\mathcal{D}$ will modulate $m_{\widetilde{A}}(x)$ to reflect this judgment by either decreasing (or increasing) it; see Section 5.4. $\mathcal{D}$'s opinion of what $\mathcal{Z}$ would say were $x \notin \widetilde{A}$ enters the picture when one is interested in evaluating the constant of proportionality in (8); see Section 5.5. Thus the membership function, when interpreted as a likelihood that encapsulates $\mathcal{D}$'s assessment of $\mathcal{Z}$'s expertise for classifying every x in $\widetilde{A}$, provides an "operational meaning" to membership functions more in the sense of Deming (1986, chap. 9) than that of de Finetti (1974) who made probability operational by interpreting it as a two-sided bet. To Deming, an operational definition puts communicable meaning into a concept, and the only communicable meaning of any word is the record of what happens on an application of a specified operation. In the context considered here, by interpreting the membership function as a likelihood, $\mathcal{D}$ is able to communicate his or her assessment of $\mathcal{Z}$'s expertise via the manner in which $\mathcal{D}$ modulates the membership function. Whereas the calculus of probability can provide a vehicle for combining likelihoods—though it need not be the only vehicle—the rules of combining membership functions are based not on probabilistic considerations but rather, as discussed in Section 3.2, on set theoretic considerations. However, this does not imply that the rules of membership functions that affect the right side of (8) create an incompatibility with the rules of probability that operate on the left side. Indeed, it is the rules of probability that operate on both sides of (8), the membership function being simply a weighting function. Consequently, the maxima and minima rules for combining fuzzy sets can coexist with the addition and multiplication rules of probability. This is because each is invoked on a different entity.

Contrast the measure given by (8) with that obtained by Zadeh (1968); see (1). Whereas Zadeh gave the expected value of the membership function as a probability measure of $\widetilde{A}$, we have argued that the probability measure of $\widetilde{A}$ is proportional to the expected value of the product of $\mathcal{Z}$'s membership function $m_{\widetilde{A}}(x)$ and $\mathcal{D}$'s probability of nature's classification of each $x \in \Omega$ in $\widetilde{A}$. In both cases, the expectations are with respect to the prior probability that $X = x$.

In principle, (8) represents our attempt at making fuzzy set theory and probability theory work in concert. However, there could be variations on the theme described earlier, one of which pertains to $\mathcal{D}$'s treatment of $m_{\widetilde{A}}(x)$ and the manner in which $\mathcal{D}$ is able to fuse several membership functions. We discuss this next. However, before doing so we find it useful to clarify a question that may arise as to whether or not $m_{\widetilde{A}}(x)$ is a surrogate for $\mathcal{Z}$'s prior probability that $x \in \widetilde{A}$, just like how $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ represents $\mathcal{D}$'s prior probability that $x \in \widetilde{A}$. Our answer is "no," because in writing $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, $\mathcal{D}$ adheres to the law of the excluded middle, whereas in specifying $m_{\widetilde{A}}(x)$, $\mathcal{Z}$ rejects the law of the excluded middle. That is, to $\mathcal{Z}$ any x need not belong to $\widetilde{A}$ to its complement, whereas to the genie an x either does or does not belong to $\widetilde{A}$.

5.4 Modulating and Pooling Membership Functions

The setup that we have introduced three agents, nature, $\mathcal{D}$, and $\mathcal{Z}$, with the last two contemplating the actions of the first and $\mathcal{D}$ in turn contemplating the actions of $\mathcal{Z}$. Advantages of this setup are the facility to modulate the input of an expert and the ability to pool the membership functions of several experts. To the best of our knowledge, neither this facility nor this mode of thinking is available in the state of the art of fuzzy set theory and in fuzzy control.

Equation (7) is based on the premise that $\mathcal{D}$ accepts $\mathcal{Z}$'s testimony as is. However, $\mathcal{D}$ may feel that $\mathcal{Z}$ may suffer from certain biases that could cause $\mathcal{Z}$ to be better (or worse) at assessing the disposition of certain values of x over the others, so that from $\mathcal{D}$'s point of view,

$$\mathcal{L}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) = m_{\widetilde{A}}(x)$$

for certain values of x, and that for the other values of x,

$$\mathcal{L}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) = m^*_{\widetilde{A}}(x),$$

where $m^*_{\widetilde{A}}(x)$ represents $\mathcal{D}$'s modulation of $m_{\widetilde{A}}(x)$ to capture the biases mentioned before. The specifics of how to arrive at an $m^*_{\widetilde{A}}(x)$ have been given by from Lindley (1983).

In a similar vein, suppose that $\mathcal{D}$ were to consult several experts, say $\mathcal{Z}_1, \ldots, \mathcal{Z}_k$, each of whom provides $\mathcal{D}$ with their respective membership functions, $m^{(1)}_{\widetilde{A}}(x), \ldots, m^{(k)}_{\widetilde{A}}(x)$, $x \in \Omega$. Then $\mathcal{D}$ may modulate each membership function as indicated earlier and then pool (or fuse) the k membership functions, incorporating, if necessary, correlations between the expert testimonies. The strategy for pooling based on the calculus of probability, was laid out by Lindley (1983) (also see Meyer and Booker 2001). The details are not given here.

5.5 The Normalized Probabilities of Fuzzy Sets

Equation (8) provides $\mathcal{D}$'s unnormalized probability measure of the fuzzy set $\widetilde{A}$. To obtain a normalized measure, we appeal to the feature that because $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ is $\mathcal{D}$'s probability that nature classifies x as being in $\widetilde{A}$, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}) + \mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A})$ should equal one; recall, that in specifying $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, $\mathcal{D}$ subscribes to the law of the excluded middle. Consequently, (5) can be written as

$$\begin{aligned}
&\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A} | m_{\widetilde{A}}(x)\big)\\
&\quad = \mathcal{P}_{\mathcal{D}}\big(m_{\widetilde{A}}(x) | x \in \widetilde{A}\big)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})\\
&\qquad \times \big(\mathcal{P}_{\mathcal{D}}\big(m_{\widetilde{A}}(x) | x \in \widetilde{A}\big)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})\\
&\qquad\qquad + \mathcal{P}_{\mathcal{D}}\big(m_{\widetilde{A}}(x) | x \notin \widetilde{A}\big)\mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A})\big)^{-1}.
\end{aligned}$$

Because $m_{\widetilde{A}}(x)$ has been specified, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} | m_{\widetilde{A}}(x))$ gets replaced by $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x))$ and $\mathcal{P}_{\mathcal{D}}(m_{\widetilde{A}}(x) | x \in \widetilde{A})$ gets replaced by $\mathcal{L}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x))$. Similarly, $\mathcal{P}_{\mathcal{D}}(m_{\widetilde{A}}(x) | x \notin \widetilde{A})$ gets replaced by $\mathcal{L}_{\mathcal{D}}(x \notin \widetilde{A}; m_{\widetilde{A}}(x))$. This last quantity is $\mathcal{D}$'s likelihood that $\mathcal{Z}$ will declare some function of $m_{\widetilde{A}}(x)$, say $1 - m_{\widetilde{A}}(x)$, as a reflection of the extent to which $x \notin \widetilde{A}$ when nature does not classify an x as belonging to $\widetilde{A}$. With the foregoing in place, we have

$$\begin{aligned}
&\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \\
&\quad = \mathcal{L}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}) \\
&\qquad \times \big(\mathcal{L}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}) \\
&\qquad\qquad + \mathcal{L}_{\mathcal{D}}\big(x \notin \widetilde{A}; m_{\widetilde{A}}(x)\big)\mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A})\big)^{-1} \\
&\quad = \left[1 + \frac{\mathcal{L}_{\mathcal{D}}(x \notin \widetilde{A}; m_{\widetilde{A}}(x))\mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A})}{\mathcal{L}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x))\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})}\right]^{-1} \\
&\quad = \left[\begin{matrix} 1 + (\text{likelihood ratio of membership functions}) \\ \times(\text{prior odds}) \end{matrix}\right]^{-1}. \qquad (9)
\end{aligned}$$

Equation (9) which is reminiscent of the well-known relationship between prior and posterior odds, is reassuring. Note that when $\widetilde{A}$ is a crisp set, $m_{\widetilde{A}}(x) = 1$ or 0 and the ratio of likelihoods is 1; thus the right side of (9) reduces to $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, as it should. Furthermore, it can be verified that when $\mathcal{L}_{\mathcal{D}}(x \in \widetilde{A}, m_{\widetilde{A}}(x)) = m_{\widetilde{A}}(x)$ and $\mathcal{L}_{\mathcal{D}}(x \notin \widetilde{A}; m_{\widetilde{A}}(x)) = 1 - m_{\widetilde{A}}(x)$, so that the event $(x \notin \widetilde{A}) = (x \in \widetilde{A}^c)$, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x)) + \mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A}; m_{\widetilde{A}}(x)) = 1$ for any and every x. Indeed, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x)) + \mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A}; m_{\widetilde{A}}(x)) = 1$, irrespective of the functional form of the likelihood. This observation reaffirms the claim of Bellman and Giertz (1973) that there is no unique way of defining $\widetilde{A}^c$, the complement of a fuzzy set. For example, one could define $\widetilde{A}^c = \{x : m_{\widetilde{A}}(x) = 0\}$; that is, the complement of a fuzzy set could be a crisp set. This is why we consider events of the type $(x \in \widetilde{A})$ and $(x \notin \widetilde{A})$ rather than the events of the type $(x \in \widetilde{A})$ and $(x \in \widetilde{A}^c)$, because the latter would entail adopting a specific definition of $\widetilde{A}^c$.

Because $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}|m_{\widetilde{A}}(x)) = \sum \mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}|m_{\widetilde{A}}(x))\mathcal{P}_{\mathcal{D}}(x)$, $\mathcal{D}$'s normalized probability measure of a fuzzy set $\widetilde{A}$ will be given as

$$\begin{aligned}
&\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \\
&\quad = \sum_x \left[1 + \frac{\mathcal{L}_{\mathcal{D}}(x \notin \widetilde{A}; m_{\widetilde{A}}(x))}{\mathcal{L}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x))} \cdot \frac{\mathcal{P}_{\mathcal{D}}(x \notin \widetilde{A})}{\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})}\right]^{-1} \\
&\qquad \times \mathcal{P}_{\mathcal{D}}(x), \qquad (10)
\end{aligned}$$

an expression that is substantially different from that given by Zadeh, namely our (1). Finally, it follows from (9) that

$$\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}; m_{\widetilde{A}}(x)\big) + \mathcal{P}_{\mathcal{D}}\big(X \notin \widetilde{A}; m_{\widetilde{A}}(x)\big) = 1.$$

5.6 Invoking the Calculus of Probability on Fuzzy Sets

We have argued that for a fuzzy set $\widetilde{A}$ with membership function $m_{\widetilde{A}}(x)$,

$$\begin{aligned}
&\mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}) \\
&\quad \equiv \mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \propto \sum_x \mathcal{P}_{\mathcal{D}}(x) m_{\widetilde{A}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}),
\end{aligned}$$

and that $\mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}) + \mathcal{P}_{\mathcal{D}}(X \notin \widetilde{A}) = 1$; the event $(X \notin \widetilde{A})$ is the complement of the event $(X \in \widetilde{A})$. In particular, if $\widetilde{A}^c$ is the fuzzy set whose membership function is $(1 - m_{\widetilde{A}}(x))$, then $\mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}) + \mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}^c) = 1$.

The purpose of this section is to argue that when a measure of the type $\mathcal{P}_{\mathcal{D}}(\widetilde{A})$ is invoked on collections of fuzzy sets for the purpose of combining them, the calculus of probability is upheld. In particular, the subset property and the (finite) additivity property of probability measures continues to hold. Furthermore, we are able to define the independence property of fuzzy sets, which enables us to prescribe a multiplication rule for fuzzy sets. To do this, we first need to make precise the notion of disjoint fuzzy sets. We start by recalling that $\widetilde{A} \cap \widetilde{B}$ is a fuzzy set whose membership function is $m_{\widetilde{A}\cap\widetilde{B}}(x) = \min(m_{\widetilde{A}}(x), m_{\widetilde{B}}(x))$. Consequently, $\widetilde{A}$ and $\widetilde{B}$ are disjoint if $\min(m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)) = 0\ \forall x \in \Omega$. We thus have the following.

Claim 1. If $\widetilde{A}$ and $\widetilde{B}$ are disjoint, then $\mathcal{P}_{\mathcal{D}}(\widetilde{A} \cap \widetilde{B}) = 0$.

To assert the subset property, we have the following.

Claim 2. If $\widetilde{B} \subseteq \widetilde{A}$, then $\mathcal{P}_{\mathcal{D}}(\widetilde{B}) \le \mathcal{P}_{\mathcal{D}}(\widetilde{A})$.

Proof. If $\widetilde{B} \subseteq \widetilde{A}$, then, according to $\mathcal{Z}$, $m_{\widetilde{B}}(x) \le m_{\widetilde{A}}(x)$ $\forall x \in \Omega$.

Now, for every $x \in \Omega$,

$$\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \propto \mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}) m_{\widetilde{A}}(x)$$

and

$$\mathcal{P}_{\mathcal{D}}\big(x \in \widetilde{B}; m_{\widetilde{B}}(x)\big) \propto \mathcal{P}_{\mathcal{D}}(x \in \widetilde{B}) m_{\widetilde{B}}(x).$$

Because $m_{\widetilde{B}}(x) \le m_{\widetilde{A}}(x)$, and because nature upholds the law of the excluded middle, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{B}) \le \mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$ for all $x \in \Omega$, assuming that $\mathcal{D}$ is in agreement with $\mathcal{Z}$ that $\widetilde{B} \subseteq \widetilde{A}$. It now follows—after some algebra—that $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}; m_{\widetilde{A}}(x)) \ge \mathcal{P}_{\mathcal{D}}(x \in \widetilde{B}; m_{\widetilde{B}}(x))$ for each x, so that $\mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}) \ge \mathcal{P}_{\mathcal{D}}(X \in \widetilde{B})$.

The countable additivity property of disjoint fuzzy sets is asserted by the following.

Claim 3. If $\widetilde{A}$ and $\widetilde{B}$ are disjoint, in the sense described earlier, then

$$\mathcal{P}_{\mathcal{D}}(X \in \widetilde{A} \cup \widetilde{B}) = \mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}) + \mathcal{P}_{\mathcal{D}}(X \in \widetilde{B}).$$

Proof. Because $\widetilde{A}$ and $\widetilde{B}$ are disjoint, there exists a crisp set $\mathcal{A}$ such that, for all $x \in \mathcal{A}$, $m_{\widetilde{A}}(x) \ge m_{\widetilde{B}}(x)$, and vice versa for all $x \in \mathcal{A}^c$. Thus

$$\begin{aligned}
&\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A} \cup \widetilde{B}; m_{\widetilde{A}\cup\widetilde{B}}(x)\big) \\
&\quad \propto \sum_x \mathcal{P}_{\mathcal{D}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} \cup \widetilde{B}) m_{\widetilde{A}\cup\widetilde{B}}(x) \\
&\quad = \sum_x \mathcal{P}_{\mathcal{D}}(x)\big[\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}) + \mathcal{P}_{\mathcal{D}}(x \in \widetilde{B})\big] \\
&\qquad\qquad \times \max\big(m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)\big) \\
&\quad = \sum_{x\in\mathcal{A}} \mathcal{P}_{\mathcal{D}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A}) m_{\widetilde{A}}(x) \\
&\qquad\quad + \sum_{x\in\mathcal{A}^c} \mathcal{P}_{\mathcal{D}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{B}) m_{\widetilde{B}}(x),
\end{aligned}$$

from which the result follows.

In the foregoing proof, the first equality is a consequence of the additivity property of probability for disjoint crisp sets. To specify a multiplication rule for fuzzy sets, we need to articulate the notion of a *conditional probability* for fuzzy sets $\widetilde{A}$ and $\widetilde{B}$. We denote this by $\mathcal{P}_{\mathcal{D}}(\widetilde{A}|\widetilde{B}) = \mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}|X \in \widetilde{B}; m_{\widetilde{A}}(x), m_{\widetilde{B}}(x))$, and following the standard convention, define it as

$$\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A} \cap \widetilde{B}; m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)\big)/\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{B}; m_{\widetilde{B}}(x)\big),$$

where the numerator term is proportional to

$$\sum_x \mathcal{P}_{\mathcal{D}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} \cap \widetilde{B}) \min\big(m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)\big)$$

and the denominator term is proportional to

$$\sum_x \mathcal{P}_{\mathcal{D}}(x)\mathcal{P}_{\mathcal{D}}(x \in \widetilde{B})m_{\widetilde{B}}(x).$$

Finally, as a consequence of the foregoing, we have the notion of independent fuzzy sets. Ignoring the presence of a conditioning event (which is required for an all encompassing definition of independence), we have the following definition.

Definition 1. The event $(X \in \widetilde{A})$ is independent of the event $(X \in \widetilde{B})$, denoted by $\widetilde{A} \perp \widetilde{B}$, if

$$\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A} \cap \widetilde{B}; m_{\widetilde{A}\cap\widetilde{B}}(x)\big) = \mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}; m_{\widetilde{A}}(x)\big)\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{B}; m_{\widetilde{B}}(x)\big)$$

or, equivalently, if

$$\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}|X \in \widetilde{B}; m_{\widetilde{A}}(x), m_{\widetilde{B}}(x)\big) = \mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}; m_{\widetilde{A}}(x)\big),$$

and vice versa when we condition on $(X \in \widetilde{A})$.

Intuitively, one would expect that if $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} \cap \widetilde{B}) = \mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})\mathcal{P}_{\mathcal{D}}(x \in \widetilde{B})$ for all x, then $\widetilde{A} \perp \widetilde{B}$. However, this need not be true, because the operation of taking $\min(m_{\widetilde{A}}(x), m_{\widetilde{B}}(x))$ could also induce dependence and destroy the independence implicit in $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A} \cap \widetilde{B})$.

6. DECISION MAKING IN A FUZZY ENVIRONMENT

In Section 3.4 we alluded to the fact that fuzzy set theory appears to have had its biggest success in the context of control theory when the actions, outcomes, or consequences cannot be precisely delineated. To illustrate the foregoing, we consider a simple example of decision making under uncertainty wherein the outcome (state) space cannot be precisely specified. The example, although simplistic, can occur in real life decision making. A more substantive example discussing fuzzy logic controllers and a probabilistic alternative to these has been given by Barrett and Woodall (1997); these authors also provided a nice overview of the essence of control based on fuzzy logic.

Suppose that the state space pertains to the weather conditions at a location to be visited on a particular future date. Let X denote the average temperature for the date in question. Because X is unknown, let $P_{\mathcal{D}}(x)$ denote $\mathcal{D}$'s personal probability that $X = x$; $\mathcal{D}$ is a decision maker.

Now it seems reasonable to partition the state space into three fuzzy subsets, with $\widetilde{A}$ denoting a warm day, $\widetilde{B}$ denoting a cool day, and $\widetilde{C}$ denoting a cold day. Associated with the foregoing, $\mathcal{D}$ may also choose to elicit three membership functions, $m_{\widetilde{A}}(x)$, $m_{\widetilde{B}}(x)$, and $m_{\widetilde{C}}(x)$, and, for each x, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{A})$, $\mathcal{D}$'s personal probability that x is classified (by nature) in $\widetilde{A}$; similarly, $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{B})$ and $\mathcal{P}_{\mathcal{D}}(x \in \widetilde{C})$. With the foregoing in place, $\mathcal{D}$ will be able to assess $\mathcal{P}_{\mathcal{D}}(X \in \widetilde{A}; m_{\widetilde{A}}(x))$ via (10). Similarly, $\mathcal{D}$ will be able to assess $\mathcal{P}_{\mathcal{D}}(X \in \widetilde{B}; m_{\widetilde{B}}(x))$ and $\mathcal{P}_{\mathcal{D}}(X \in \widetilde{C}; m_{\widetilde{C}}(x))$.

$\mathcal{D}$'s interest in the weather has to do with having to make a decision about packing clothing for the contemplated visit. Suppose that $\mathcal{D}$'s action (decision) space is limited to the following three crisp choices: TC, take a topcoat; S, take a sweater; or N, take neither. The decision tree shown in Figure 2 portrays the elements of the action space, subsequent to the decision node (the rectangle), and the state space, subsequent to the random node (the circle). Barrett and Woodall (1997) considered the control of an air-conditioner wherein the action space is also paritioned into fuzzy sets. The terminus of the tree shows $\mathcal{D}$'s utilities for each action–state combination. For example, $\mathcal{U}_{\mathcal{D}}(N, \widetilde{C})$ is $\mathcal{D}$'s utility (actually disutility) in taking neither a topcoat nor a pullover on a cold day. For coherent decision making, $\mathcal{D}$ needs

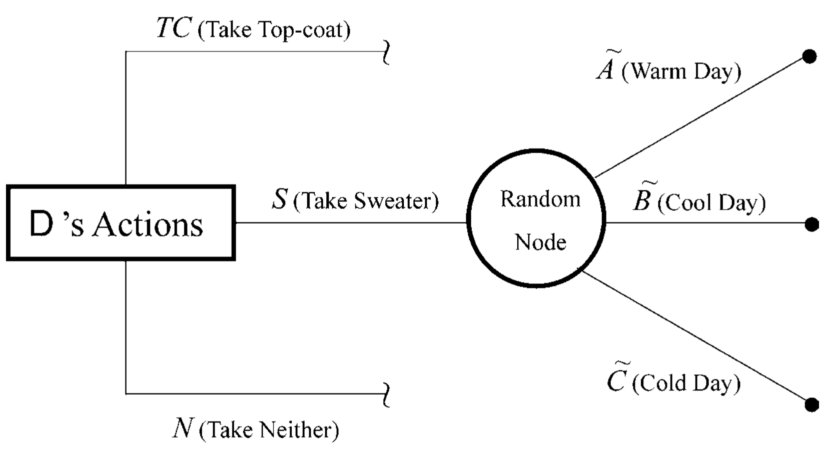

Figure 2. Decision Tree for Fuzzy Outcomes.

to specify his or her utilities for each action–state combination. For convenience, Figure 2 shows these only for the case of the action designated by S.

Following the principle of maximization of expected utility (MEU) (Lindley 1991, p. 58), $\mathcal{D}$ should choose that action for which the expected utility is a maximum where, for example, the expected utility should $\mathcal{D}$ choose action S is of the form

$$\begin{aligned}\mathbf{E}[\mathcal{U}_{\mathcal{D}}(S)] = {} & \mathcal{U}_{\mathcal{D}}(S, \widetilde{A})\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{A}; m_{\widetilde{A}}(x)\big) \\ & + \mathcal{U}_{\mathcal{D}}(S, \widetilde{B})\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{B}; m_{\widetilde{B}}(x)\big) \\ & + \mathcal{U}_{\mathcal{D}}(S, \widetilde{C})\mathcal{P}_{\mathcal{D}}\big(X \in \widetilde{C}; m_{\widetilde{C}}(x)\big).\end{aligned}$$

The foregoing example may seem uninteresting at first glance; however, its control theory version, which could be a climate control system involving three fan speed actions, (low, medium, and high) is quite realistic. A thermometer could anticipate the temperature as being normal, warm, or hot, and based on this, the control system could trigger a fan speed. Other such examples may be constructed similarly.

It is unlikely that control theorists working on problems labeled "fuzzy control" would follow the prescription outlined here; to them, considering fuzzy sets would be tantamount to discarding the calculus of probability and thus the principle of maximization of expected utility. Consequently, they have to resort to several ad hoc rules known as "correlation minimum encoding" and "correlation product encoding" (Barrett and Woodall 1997). The purpose of this article is to argue that this need not be so. Fuzzy control can be achieved within the framework of probability, even if membership functions are not interpreted as (conditional) probabilities, which Barrett and Woodall (1997) did.

7. CONCLUDING PERSPECTIVE

It has been pointed out by a referee—and correctly so—that statisticians need to be more interdisciplinary in their outlook. Hopefully, this article reflects the spirit of this sentiment. There is a large community of engineers and computer scientists interested in the various approaches to treating uncertainty. Most of them subscribe to probability and its calculus. However, some have raised questions about its restrictions (e.g., adherence to the law of the excluded middle) and thus have searched for alternatives to probability. Possibility theory, one such alternative, has a large following, especially among those who design engineering controls. In abandoning probability, such control theorists have resorted to decision rules whose foundations are not transparent. In this article we have argued that it is possible to endow probability measures on fuzzy sets. Thus it is possible to do fuzzy control (de facto decision making under uncertainty) within the calculus of probability via the principle of maximization of expected utility. While endowing probability measures on fuzzy sets by interpreting the membership function as a likelihood, we have uncovered a feature that is not available within the calculus of fuzzy set theory and fuzzy control—specifically, we can use Bayes's law to pool several membership functions. Furthermore, given that membership functions are subjectively specified, we can modulate such functions to incorporate biases, interdependencies, and so on. The concepts of modulating and pooling membership functions appear to be alien to the community of fuzzy control. These, plus an ability to stay within the umbrella of probability and statistics, are some of the advantages of our approach.

Another prominent community interested in the treatment of uncertainty is a subfield of the field of "artificial intelligence." Their annual conference on Uncertainty in Artificial Intelligence (UAI) is a celebrated event. Fortunately, many in this arena tend to work within probability and statistics, although not always within the fully subjectivistic Bayesian paradigm. A pulse of what is going on in UAI can be gauged by reviewing a recent article by Giang and Shenoy (2002). Like us, these authors focused on the likelihood function (and the "likelihood principle"), but unlike us, they forsook the prior. They argued that the likelihood function is basically a "possibility measure," and in so doing they could be paving the way for a more cogent discussion of possibility theory. In contrast, by incorporating both the likelihood and the prior, our development stays within the calculus of probability. Giang and Shenoy (2001) explored the axiomatics of qualitative decision making based on possibility theory. Our view is that if the raison d'etre of possibility theory is an acknowledgment of the need for fuzzy sets, then an ability to endow probabilities on the latter enables decision making via the MEU principle.

Some other approaches for the treatment of uncertainty are the Spohnian and the Dempster-Schaferian (Dempster 1967; Giang and Shenoy 2000; Shafer 1976; Walley 1996; Wasserman 1990). In many respects these approaches germinate from the foundations of probability theory, and like what we have done herein, they enhance probability, not supplant it. For example, Dempster's (1967) upper and lower probabilities come into being because only a many-to-one map can induce probabilities from a probability measure space to some other space, such as the space of real numbers.

The quantification of uncertainty is poised to be one of the key sciences of this era. Our hope is that the calculus of probability, with its generalizations and enhancements to encompass such notions as causality (Lindley 2002; Pearl 2000; Singpurwalla 2002), set valued maps (Dempster 1968), and vagueness (Zadeh 1986a), will be able to stand the test of time.

[Received July 2003. Revised February 2004.]

REFERENCES

Barrett, J. D., and Woodall, W. H. (1997), "A Probabilistic Alternative to Fuzzy Logic Controllers," *IIE Transactions*, 29, 459–467.

Basu, D. (1975), "Statistical Information and Likelihood," *Sankhyā*, Ser. A, 37, 1–71.

Bellman, R., and Giertz, M. (1973), "On the Analytic Formalism of the Theory of Fuzzy Sets," *Information Sciences*, 5, 149–156.

de Finetti, B. (1974), *Theory of Probability*, New York: Wiley.

Deming, W. E. (1986), *Out of the Crisis*, Cambridge, MA: MIT Press.

Dempster, A. P. (1967), "Upper and Lower Probabilities Induced by a Multivalued Mapping," *The Annals of Mathematical Statistics*, 39, 325–339.

——— (1968), "Upper and Lower Probabilities Generated by a Random Closed Interval," *The Annals of Mathematical Statistics*, 39, 957–966.

Giang, P. H., and Shenoy, P. P. (2000), "A Qualitative Linear Utility Theory for Spohn's Theory of Epistemic Beliefs," in *Uncertainty in Artificial Intelligence: Proceedings of the Sixteenth Conference*, San Fransisco: Morgan Kauffman.

——— (2001), "A Comparison of Axiomatic Approaches to Qualitative Decision Making Using Possibility Theory," in *Uncertainty in Artificial Intelligence: Proceedings of the Seventeenth Conference*, San Fransisco: Morgan Kauffman.

——— (2002), "Statistical Decisions Using Likelihood Information Without Prior Probabilities," in *Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference*, San Fransisco: Morgan Kauffman.

Gigerenzer, G. Swijtink, Z., Porter, T., Daston, L., Beatly, J., and Kruger, L. (1990), *The Empire of Chance*, Cambridge, U.K.: Cambridge University Press.

Kosko, B., and Isaka, S. (1993), "Fuzzy Logic," *Scientific American*, 269, 76–81.

Laviolette, M., Seaman, J. W., Barrett, J. D., and Woodall, W. H. (1995), "A Probabilistic and Statistical View of Fuzzy Methods," *Technometrics*, 37, 249–261.

Lindley, D. V. (1983), "Reconciliation of Probability Distributions," *Operations Research*, 31, 866–880.

——— (1987), "The Probability Approach to the Treatment of Uncertainty in Artifical Intelligence and Expert Systems," *Statistical Science*, 2, 17–24.

——— (1991), *Making Decisions*, New York: Wiley.

——— (2002), "Seeing and Doing: The Concept of Causation," *International Statistical Review*, 70, 191–214.

Loginov, V. I. (1966), "Probability Treatment of Zadeh Membership Functions and Their Use in Pattern Recognition," *Engineering Cybernetics*, 68–69.

Meyer, M., and Booker, J. (2001), *Eliciting and Analyzing Expert Judgment—A Practical Guide*, Philadelphia: ASA-SIAM Publications.

Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, New York: Cambridge University Press.

Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton, NJ: Princeton University Press.

Singpurwalla, N. D. (2002), "On Causality and Causal Mechanisms," *International Statistical Review*, 70, 198–214.

Walley, P. (1996), "Measures of Uncertainty in Expert Systems," *Artificial Intelligence*, 83, 1–58.

Wasserman, L. A. (1990), "Belief Functions and Statistical Inference," *The Canadian Journal of Statistics*, 18, 183–196.

Zadeh, L. A. (1965), "Fuzzy Sets," *Information and Control*, 8, 338–353.

——— (1968), "Probability Measures of Fuzzy Events," *Journal of Mathematical Analysis and Applications*, 23, 421–427.

——— (1978), "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems*, 1, 3–28.

——— (1986a), "Is Probability Theory Sufficient for Dealing With Uncertainty in AI?: A Negative View," in *Uncertainty in Artificial Intelligence*, eds. L. N. Kanal and J. F. Lemmer, Amsterdam: North-Holland, pp. 103–106.

——— (1986b), "A Simple View of the Dempster–Shafer Theory of Evidence and Its Implications for the Rule of Combination of Evidence," *AI Magazine*, 7, 85–90.

——— (1995), "Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather Than Competitive," *Technometrics*, 37, 271–276.

Comment

D. V. LINDLEY

One of the difficulties many people have with the ideas of fuzzy logic lies in the interpretation of the membership function: What does it mean to say that $m(x) = .2$? [The subscript denoting the set, used by Singpurwalla and Booker (SB), is omitted here because only one fuzzy set is being discussed.] Suppose that you were interested in a quantity, uncertain for you; then a modified form of the question would be to ask how, being told the values of a membership function, your uncertainty of the quantity would be changed. In probability terms, how would $p(y)$, your probability for the quantity y, be changed to $p(y|m(\cdot))$ on being provided with the membership function $m(\cdot)$? [It is helpful in what follows to distinguish between the function $m(\cdot)$ and $m(x)$, its value at x.] SB provide an answer to this modified question. Sometimes posing the question sensibly is as important as answering it, and SB are to be congratulated on putting the question this way. The question is brilliant, but I have some reservations about the answer. To explain these, let me work through their argument using a slightly different notation, which will hopefully clarify my qualifications.

Their thesis introduces two people: $\mathcal{Z}$, who supplies $m(\cdot)$, and $\mathcal{D}$ (the "you" in the preceeding paragraph), who is uncertain. Because all probabilities belong to $\mathcal{D}$, we also omit that subscript from all probabilities. An important introduction by SB is the genie who sees a sharp set A, not a fuzzy one, but never reveals the sharp set's membership function, so that $\mathcal{Z}$ and $\mathcal{D}$ remain fuzzy in their understanding. As SB wisely say in Section 5.1, "Fuzzy sets are only a manifestation of *our* uncertainty about ... sharp sets." The membership function of A at x is written as $\theta(x)$, because it plays the familiar role of a parameter in statistical inference, where the Greek alphabet is commonly used, and assumes only one of two values, 0 or 1. $\mathcal{D}$ then has $p(x)$, the usual probability for x, and $p(\theta(x)|x)$, the probability that x does or does not belong to A, giving

$$p(\theta) = \sum_x p\big(\theta(x)|x\big)p(x). \qquad [2]\,(1)$$

(The numbers in square brackets refer to the corresponding equation in SB.) $p(\theta)$, on the left side of (1), is $\mathcal{D}$'s probability that a random x belongs ($\theta = 1$) or does not belong ($\theta = 0$) to A. In (1) and what follows we write θ, in lieu of $\theta(x)$, whenever the conditions include x, the argument being that if x is known, then only $\theta(x)$ is relevant. $\theta(\cdot)$, the complete membership function of sharp A, does not enter into our calculations.

SB's splendid idea is to see how beliefs about A change on $\mathcal{D}$ being supplied with a membership function by $\mathcal{Z}$. Before we follow their trail, let us consider the simpler case where $\mathcal{D}$ is interested in a hypothesis H and $\mathcal{Z}$ says "yes," Y, when asked whether it is true. Then $\mathcal{D}$'s odds on H will change to

$$O(H|Y) = p(Y|H)/p(Y|H^c) \cdot O(H) \qquad (2)$$

by Bayes's rule. The key point to notice is that $\mathcal{D}$ will need to assess two probabilities, or at least their ratio: the probability that $\mathcal{Z}$ will respond with Y both when it is true and when H is false, H^c. With this consideration in mind, let us return to fuzzy logic, where we need to evaluate $p(\theta|m(\cdot))$, the change from $p(\theta)$ when $\mathcal{Z}$ supplies $m(\cdot)$. SB extend the conversation to include x, as in (1), giving

$$p\big(\theta|m(\cdot)\big) = \sum_x p\big(\theta|m(\cdot), x\big)p\big(x|m(\cdot)\big), \qquad [3]\,(3)$$

D. V. Lindley was formerly Professor, University College, London; he is now retired and lives at Woodstock, Quay Lane, Minehead, Somerset TA 24 5QU, England.

Journal of the American Statistical Association
September 2004, Vol. 99, No. 467, Review Article
DOI 10.1198/016214504000001204