Foundations of Physics, Vol. 30, No. 10, 2000

What Is Fuzzy Probability Theory?

S. Gudder[1]

Received March 4, 1998; revised July 6, 2000

The article begins with a discussion of sets and fuzzy sets. It is observed that identifying a set with its indicator function makes it clear that a fuzzy set is a direct and natural generalization of a set. Making this identification also provides simplified proofs of various relationships between sets. Connectives for fuzzy sets that generalize those for sets are defined. The fundamentals of ordinary probability theory are reviewed and these ideas are used to motivate fuzzy probability theory. Observables (fuzzy random variables) and their distributions are defined. Some applications of fuzzy probability theory to quantum mechanics and computer science are briefly considered.

1. INTRODUCTION

What do we mean by fuzzy probability theory? Isn't probability theory already fuzzy? That is, probability theory does not give precise answers but only probabilities. The imprecision in probability theory comes from our incomplete knowledge of the system but the random variables (measurements) still have precise values. For example, when we flip a coin we have only a partial knowledge about the physical structure of the coin and the initial conditions of the flip. If our knowledge about the coin were complete, we could predict exactly whether the coin lands heads or tails. However, we still assume that after the coin lands, we can tell precisely whether it is heads or tails. In fuzzy probability theory, we also have an imprecision in our measurements, and random variables must be replaced by fuzzy random variables and events by fuzzy events.

Since fuzzy events are essentially fuzzy sets, we begin with a comparison of sets and fuzzy sets. This comparison is made evident by identifying a set

[1] Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208.

0015-9018/00/1000-1663$18.00/0

with its indicator function. Making this identification also provides simplified proofs of various relationships between sets. Connectives for fuzzy sets that generalize those for sets are defined. In particular, we define complements, intersections, unions, orthogonal sums, differences and symmetric differences for fuzzy sets. We then study some of the properties of these fuzzy connectives.

The fundamentals of ordinary probability theory are reviewed and these ideas are used to motivate fuzzy probability theory. Effects (fuzzy events), observables (fuzzy random variables) and their distributions are defined. It is shown that a finite set of observables always possesses a joint distribution.

Some applications of fuzzy probability theory to quantum mechanics and computer science are briefly considered. It is noted that the set of effects $\mathcal{E}$ for a fixed system forms a σ-effect algebra and these algebras have recently been important in studies of the foundations of quantum mechanics. It is shown that there exists a natural bijection between the set of states on $\mathcal{E}$ and the set of probability measures on the underlying sample space. Moreover, there is a natural one-to-one correspondence between σ-morphisms of these effect algebras and observables. This correspondence enables us to define a composition of observables. Some of these ideas are then applied in a discussion of indeterministic automata.

It is the intention of this article to give a brief survey of the subject. For more details and alternative approaches, we refer the reader to the literature.[1, 2, 4–6, 10, 13–15, 17] We congratulate Marisa Dalla Chiara on this special occasion and take great pleasure in acknowledging her contributions and influence. She is a guiding spirit, and we treasure her presence.

2. SETS AND FUZZY SETS

Let Ω be a nonempty set and let 2^Ω be its power set. Corresponding to any $A \in 2^\Omega$ we define its *indicator function* I_A by

$$I_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A \end{cases}$$

We can identify A with I_A because $A = B$ if and only if $I_A = I_B$ and in the sequel we shall frequently treat A and I_A as the same object. Notice that $I_{A\cap B} = I_A I_B$, $I_{A\cup B} = I_A + I_B - I_A I_B$ and that $I_\varnothing = 0$, $I_\Omega = 1$ where 0 and 1 are the constant zero and one functions, respectively. It is also useful to note that the idempotent law, $I_A^2 = I_A$, holds. Denoting the complement of A by A', we have $I_{A'} = 1 - I_A$. Observe that $A \cap B = \varnothing$ if and only if

$I_A \leqslant I_{B'}$. This condition is equivalent to $I_A + I_B \leqslant 1$ and in this case we have $I_{A \cup B} = I_A + I_B$.

The identification $A \leftrightarrow I_A$ is not only useful for discussing fuzzy sets, it has advantages in ordinary set theory. For example, proving the distributive law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

using sets is a bit of trouble. However, using indicator functions we have

$$\begin{aligned} I_{A \cap (B \cup C)} &= I_A I_{B \cup C} = I_A (I_B + I_C - I_B I_C) = I_A I_B + I_A I_C - I_A I_B I_C \\ &= I_{A \cap B} + I_{B \cap C} - I_{A \cap B} I_{A \cap C} = I_{(A \cap B) \cup (A \cap C)} \end{aligned}$$

As another example, we can prove De Morgan's law $(A \cup B)' = A' \cap B'$ as follows

$$\begin{aligned} I_{A' \cap B'} &= I_{A'} I_{B'} = (1 - I_A)(1 - I_B) = 1 - (I_A + I_B - I_A I_B) \\ &= 1 - I_{A \cup B} = I_{(A \cup B)'} \end{aligned}$$

The next result is the inclusion–exclusion law.

Lemma 2.1. If $A_i \in 2^{\Omega}$, $i = 1, \ldots, n$, then

$$I_{\cup A_i} = \sum_i I_{A_i} - \sum_{i<j} I_{A_i} I_{A_j} + \sum_{i<j<k} I_{A_i} I_{A_j} I_{A_k} - \cdots + (-1)^{n-1} I_{A_1} I_{A_2} \cdots I_{A_n}$$

Proof. The result clearly holds for $n = 1$. Proceeding by induction, suppose the result holds for the integer $n \geqslant 1$. Letting $B = \bigcup_{i=1}^{n} A_i$, we have

$$I_{B \cup A_{n+1}} = I_B + I_{A_{n+1}} - I_B I_{A_{n+1}}$$

But the right side of this equation gives the result for $n + 1$. □

It is well known that 2^{Ω} is a Boolean ring under the operations $A \cdot B = A \cap B$, $A + B = (A \cap B') \cup (B \cap A')$. However, this result is very cumbersome to prove using set theoretic operations! Using indicator functions, the proof is simple and straightforward. First notice that

$$\begin{aligned} I_{A+B} &= I_{A \cap B'} + I_{B \cap A'} = I_A (1 - I_B) + I_B (1 - I_A) \\ &= I_A + I_B - 2 I_A I_B = (I_A - I_B)^2 \end{aligned}$$

Theorem 2.2. Under the previously defined operations, 2^{Ω} is a commutative, idempotent ring with identity and characteristic 2.

Proof. It is clear that 0, 1 are the zero and identity for 2^{Ω}, that A' is the additive inverse of A and that 2^{Ω} is commutative. Moreover, it is clear that $A+B=B+A$, $A \cdot A=A$, $A+A=0$ and that multiplication is associative. For associativity of addition, we have

$$\begin{aligned} I_{(A+B)+C} &= I_{A+B}+I_C-2I_{A+B}I_C \\ &= I_A+I_B-2I_AI_B+I_C-2(I_A+I_B-2I_AI_B)\,I_C \\ &= I_A+I_B+I_C-2(I_AI_B+I_AI_C+I_BI_C)+4I_AI_BI_C \\ &= I_{A+(B+C)} \end{aligned}$$

For distributivity, we have

$$\begin{aligned} I_{(A\cdot B+A\cdot C)} &= I_{A\cdot B}+I_{A\cdot C}-2I_{A\cdot B}I_{A\cdot C} \\ &= I_AI_B+I_AI_C-2I_AI_BI_C \\ &= I_A(I_B+I_C-2I_BI_C) \\ &= I_AI_{B+C}=I_{A\cdot(B+C)} \end{aligned}$$

□

We also have the following result whose proof is similar to that of Lemma 2.1.

Lemma 2.3. If $A_i \in 2^{\Omega}$, $i=1,\dots,n$, then

$$I_{\Sigma A_i}=\sum_i I_{A_i}-2\sum_{i<j} I_{A_i}I_{A_j}+2^2\sum_{i<j<k} I_{A_i}I_{A_j}I_{A_k}-\cdots+(-2)^{n-1}\,I_{A_1}I_{A_2}\cdots I_{A_n}$$

Fuzzy set theory was introduced by Zadeh[16, 17] to describe situations with unsharp boundaries, partial information or vagueness such as in natural language. In fuzzy set theory, subsets of Ω are replaced by functions $f\colon \Omega \to [0,1]$. We thus replace the power set 2^{Ω} by the function space $[0,1]^{\Omega}$. Elements of $[0,1]^{\Omega}$ are called *fuzzy sets* and an $f \in [0,1]^{\Omega}$ corresponds to a degree of membership function. We say that f is *crisp* if the values of f are contained in $\{0,1\}$. Thus, f is crisp if and only if f is an indicator function or equivalently a set in 2^{Ω}. We thus see that a fuzzy set is a generalization of a set. For $f, g \in [0,1]^{\Omega}$, if $f+g \in [0,1]^{\Omega}$ (equivalently, $f+g \leqslant 1$), then we write $f \perp g$ and define $f \oplus g = f+g$. This *orthogonal sum* partial operation corresponds to a disjoint union for sets because $I_A \perp I_B$ if and only if $A \cap B = \varnothing$ and then $I_{A\cup B}=I_A+I_B$. Just as disjoint unions are important in probability theory, orthogonal sums will be important in fuzzy probability theory.

We now define connectives for fuzzy sets that generalize those for sets. For $f, g \in [0,1]^\Omega$ we define $f' = 1 - f$, $f \cap g = fg$ and $f \cup g = f + g - fg$. These definitions correspond to the usual properties of indicator functions and we have $I'_A = I_{A'}$, $I_A \cap I_B = I_{A \cap B}$, $I_A \cup I_B = I_{A \cup B}$. It is clear that $f', f \cap g \in [0,1]^\Omega$ and also $f \cup g \in [0,1]^\Omega$ because De Morgan's law

$$f \cup g = 1 - (1-f)(1-g) = (f' \cap g')'$$

holds. We can write this as $(f \cup g)' = f' \cap g'$ and we also have $(f \cap g)' = f' \cup g'$. Notice that $f \cap g = 0$ implies $f \perp g$ but that $f \perp g$ need not imply $f \cap g = 0$. For example, if f is the constant function $f = 1/2$, then $f \perp f$ but $f \cap f \neq 0$. This is different than in set theory where $I_A \cap I_B = 0$ if and only if $I_A \perp I_B$. It is easy to show that f is crisp if and only if $f \cap f' = 0$ (equivalently, $f \cup f' = 1$). In fact, in set theory we always have $I_A \cap I'_A = I_{A \cap A'} = 0$. The following inclusion–exclusion law has the same proof as Lemma 2.1.

Lemma 2.4. If $f_i \in [0,1]^\Omega$, $i = 1, \ldots, n$, then

$$f_1 \cup \cdots \cup f_n = \sum_i f_i - \sum_{i<j} f_i f_j + \sum_{i<j<k} f_i f_j f_k - \cdots + (-1)^{n-1} f_1 f_2 \cdots f_n$$

It is interesting to examine which properties of the Boolean ring 2^Ω carry over to $[0,1]^\Omega$. For this purpose, we use the notation $f \cdot g = fg = f \cap g$ and define

$$f \boxplus g = f + g - 2fg$$

Notice that $f \boxplus g$ generalizes the definition of $A + B$ for sets and we have $I_A \boxplus I_B = I_{A+B}$. We also have

$$f \boxplus g = f(1-g) + g(1-f) = f \cap g' + g \cap f' = f \cup g - f \cap g$$

Note however, that $f \boxplus g \neq (f \cap g') \cup (g \cap f')$ in general. Now some of the properties of the Boolean ring 2^Ω hold for $([0,1]^\Omega, \cdot, \boxplus)$. Clearly, 0, 1 are the zero and identity, the operations $\cdot$, $\boxplus$ are commutative, and $\cdot$ is associative. Moreover, $\boxplus$ is associative because

$$(f \boxplus g) \boxplus h = f + g + h - 2fg - 2fh - 2gh + 4fgh = f \boxplus (g \boxplus h)$$

and we have the following analog of Lemma 2.3.

Lemma 2.5. If $f_i \in [0, 1]^{\Omega}$, $i = 1,..., n$, then

$$\boxplus f_i = \sum_i f_i - 2 \sum_{i<j} f_i f_j + 2^2 \sum_{i<j<k} f_i f_j f_k - \cdots + (-2)^{n-1} f_1 f_2 \cdots f_n$$

The next lemma shows that the other properties of 2^{Ω} do not hold. In particular, $([0, 1]^{\Omega}, \cdot, \boxplus)$ is not a ring because distributivity fails in general.

Lemma 2.6. For $f \in [0, 1]^{\Omega}$, the following statements are equivalent. (i) f is crisp, (ii) $f \cdot f = f$, (iii) $f \boxplus f = 0$, (iv) there exists a $g \in [0, 1]^{\Omega}$ such that $f \boxplus g = 1$, (v) $f \cdot (g \boxplus h) = f \cdot g \boxplus f \cdot h$ for every $g, h \in [0, 1]^{\Omega}$, (vi) $f \boxplus g = (f \cap g') \cup (g \cap f')$ for every $g \in [0, 1]^{\Omega}$.

Proof. If f is crisp, then clearly (ii)–(vi) hold. Conversely, it is obvious that (ii) and (iii) both imply that f is crisp. Now suppose that (iv) holds. It follows that $(1-g)(1-f) = -fg$. Assume that $f(w) \neq 0, 1$ for some $\omega \in \Omega$. Then

$$[1 - g(\omega)][1 - f(\omega)] = -f(\omega)\, g(\omega)$$

and the left side is positive while the right side is negative. This is a contradiction, so $f(\omega) \in \{0, 1\}$ and f is crisp. If (v) holds, then letting $g = h = 1$ we have

$$f \boxplus f = f \cdot 1 \boxplus f \cdot 1 = f \cdot (1 \boxplus 1) = f \cdot 0 = 0$$

so f is crisp. If (vi) holds, then letting $g = f$ we have

$$2ff' = 2f - 2f^2 = f \boxplus f = (f \cap f') \cup (f \cap f') = 2ff' - 2f^2(f')^2$$

Hence, $f \cap f' = 0$ so f is crisp. □

Another interesting connective for fuzzy sets is the difference operation $f \backslash g = fg'$. This operation generalizes the difference $A \backslash B = A \cap B'$ for sets. In terms of indicator functions, we have

$$I_A \backslash I_B = I_A I_{B'} = I_{A \backslash B}$$

Also, notice that $f \boxplus g = f \backslash g + g \backslash f$.

Theorem 2.7. If $f_1,..., f_n \in [0, 1]^{\Omega}$, then

$$\sum_{i=1}^{n-1} f_i \backslash f_{i+1} \geqslant f_1 \backslash f_n$$

Proof. We proceed by induction on n. For $n=2$, we have $f_1\backslash f_2 \geqslant f_1\backslash f_2$ which is certainly true. For $n=3$, we have

$$\begin{aligned} f_1\backslash f_3 &= f_1 f'_3 = f_1(f_2 + f'_2)\, f'_3 = f_1 f_2 f'_3 + f_1 f'_2 f'_3 \\ &\leqslant f_2 f'_3 + f_1 f'_2 = f_1\backslash f_2 + f_2\backslash f_3 \end{aligned}$$

so the result holds. Now suppose the result holds for an integer $n \geqslant 3$. Then applying the case $n=3$ we have

$$\sum_{i=1}^{n} f_i\backslash f_{i+1} = \sum_{i=1}^{n-1} f_i\backslash f_{i+1} + f_n\backslash f_{n+1} \geqslant f_1\backslash f_n + f_n\backslash f_{n+1} \geqslant f_1\backslash f_{n+1}$$

Hence, the result holds by induction. □

Corollary 2.8. If $A_1, \ldots, A_n \in 2^{\Omega}$, then

$$\sum_{i=1}^{n-1} I_{A_i\backslash A_{i+1}} \geqslant I_{A_1\backslash A_n}$$

We can apply Theorem 2.7 to obtain the following "triangle inequality."

Corollary 2.9. (i) $f \boxplus g \leqslant f \boxplus h + h \boxplus g$. (ii) $I_{A+B} \leqslant I_{A+C} + I_{C+B}$.

Proof. (i) By Theorem 2.7 we have

$$f\backslash g \leqslant f\backslash h + h\backslash g$$

$$g\backslash f \leqslant g\backslash h + h\backslash f$$

Adding these inequalities and using the fact that $f \boxplus g = f\backslash g + g\backslash f$ gives the result. (ii) is a special case of (i). □

3. PROBABILITY THEORY

To better appreciate fuzzy probability theory, we first review the fundamentals of ordinary probability theory. The basic structure is a measurable space $(\Omega, \mathscr{A})$ where Ω is a sample space consisting of *outcomes* andK$\mathscr{A}$ is a σ-algebra of *events* in Ω corresponding to some probabilistic experiment. It is useful to identify an event A with its indicator function I_A as we did in Sec. 2. If μ is a probability measure on $(\Omega, \mathscr{A})$, then $\mu(A)$ is interpreted as the probability that the event A occurs. A measurable functions $f: \Omega \to \mathbb{R}$ is called a *random variable*. The *expectation* $\mu(f)$ of f is defined by $\mu(f) = \int f\, d\mu$. Denoting the Borel σ-algebra on the real line $\mathbb{R}$ by $\mathscr{B}(\mathbb{R})$, the *distribution* of f is the probability measure μ_f on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given by $\mu_f(B) = \mu(f^{-1}(B))$. We interpret $\mu_f(B)$ as the probability that f has a value in the set B. It can be shown that $\mu(f) = \int \lambda \mu_f(d\lambda)$.

Notice that $\mu(I_A) = \mu(A)$ for any $A \in \mathscr{A}$ so the identification of A with I_A carries directly over to probabilities. In particular, this identification enables us to give simple proofs of basic properties of probabilities. For example, we have

$$\begin{aligned}\mu(A \cup B) &= \mu(I_{A \cup B}) = \mu(I_A + I_B - I_{A \cap B}) \\ &= \mu(I_A) + \mu(I_B) - \mu(I_{A \cap B}) = \mu(A) + \mu(B) - \mu(A \cap B)\end{aligned}$$

More generally, applying Lemma 2.1 we obtain the inclusion–exclusion law

$$\begin{aligned}\mu(\cup A_i) = \sum_i \mu(A_i) - \sum_{i<j} \mu(A_i \cap A_j) + \sum_{i<j<k} \mu(A_i \cap A_j \cap A_k) \\ - \cdots + (-1)^{n-1} \mu(A_1 \cap \cdots \cap A_n)\end{aligned}$$

For another example, define the distance between $A, B \in \mathscr{A}$ by $\rho(A, B) = \mu(A + B)$. Following the usual practice of identifying events that coincide except for a set of measure zero, we have $\rho(A, B) = 0$ if and only if $A = B$. Moreover, it follows from Corollary 2.9(ii) that the triangle inequality

$$\rho(A, B) \leqslant \rho(A, C) + \rho(C, B)$$

holds so ρ is a metric.

We call $\mathscr{E}_c(\Omega, \mathscr{A}) = \{I_A : A \in \mathscr{A}\}$ the set of crisp effects. Of course, $\mathscr{E}_c(\Omega, \mathscr{A})$ is a Boolean ring as in Theorem 2.2. Since we are describing probability theory in terms of $\mathscr{E}_c(\Omega, \mathscr{A})$, we would also like to describe random variables in terms of $\mathscr{E}_c(\Omega, \mathscr{A})$. If $f: \Omega \to \mathbb{R}$ is a random variable, define $X_f: \mathscr{B}(\mathbb{R}) \to \mathscr{E}_c(\Omega, \mathscr{A})$ by $X_f(B) = I_{f^{-1}(B)}$. Then X_f satisfies the conditions

$$X_f(\mathbb{R}) = I_{f^{-1}(\mathbb{R})} = 1$$

and if $A_i \in \mathscr{B}(\mathbb{R})$ are mutually disjoint, then

$$X_f(\cup A_i) = I_{f^{-1}(\cup A_i)} = I_{\cup f^{-1}(A_i)} = \sum I_{f^{-1}(A_i)} = \sum X_f(A_i)$$

Conversely, if $X: \mathscr{B}(\mathbb{R}) \to \mathscr{E}_c(\Omega, \mathscr{A})$ satisfies these two conditions, then it can be shown that there exists a unique random variable $f: \Omega \to \mathbb{R}$ such that $X = X_f$. We call X_f the crisp observable corresponding to f. The distribution of f can be written

$$\mu_f(B) = \mu(f^{-1}(B)) = \mu(I_{f^{-1}(B)}) = \mu(X_f(B))$$

and we call $\mu_{X_f}(B) = \mu(X_f(B))$ the distribution of X_f. The expectation of f becomes

$$\mu(f) = \int \lambda \mu_f(d\lambda) = \int \lambda \mu(X_f(d\lambda))$$

which we also call the expectation of X_f.

Some of the most important concerns in probability theory involve the study of several random variables simultaneously. The joint distribution of random variables $f_1,\dots, f_n$ on $(\Omega, \mathscr{A})$ is the unique probability measure $\mu_{f_1,\dots, f_n}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ that satisfies

$$\mu_{f_1,\dots, f_n}(B_1 \times \cdots \times B_n) = \mu(f_1^{-1}(B_1) \cap \cdots \cap f_n^{-1}(B_n))$$

for all $B_1,\dots, B_n \in \mathscr{B}(\mathbb{R})$. The joint distribution can be described by the n-dimensional random variable $J(f_1,\dots, f_n)\colon \Omega \to \mathbb{R}^n$ given by

$$J(f_1,\dots, f_n)(\omega) = (f_1(\omega),\dots, f_n(\omega))$$

We call $J(f_1,\dots, f_n)$ the joint random variable for $f_1,\dots, f_n$. Since

$$f_1^{-1}(B_1) \cap \cdots \cap f_n^{-1}(B_n) = J(f_1,\dots, f_n)^{-1}(B_1 \times \cdots \times B_n)$$

we have

$$\mu_{f_1,\dots, f_n}(B_1 \times \cdots \times B_n) = \mu(J(f_1,\dots, f_n)^{-1}(B_1 \times \cdots \times B_n))$$

It follows that

$$\mu_{f_1,\dots, f_n}(B) = \mu(J(f_1,\dots, f_n)^{-1}(B))$$

for every $B \in \mathscr{B}(\mathbb{R}^n)$.

Letting $X_{f_1},\dots, X_{f_n}$ be the corresponding crisp observables, we define their *joint crisp observable* $J(X_{f_1},\dots, X_{f_n})\colon \mathscr{B}(\mathbb{R}^n) \to \mathscr{E}_c(\Omega, \mathscr{A})$ to be the unique map that satisfies

$$J(X_{f_1},\dots, X_{f_n})(B_1 \times \cdots \times B_n) = X_{f_1}(B_1) \cdots X_{f_n}(B_n)$$

for every $B_1,..., B_n \in \mathscr{B}(\mathbb{R})$. We then have

$$J(X_{f_1},..., X_{f_n})(B_1 \times \cdots \times B_n) = I_{f_1^{-1}(B_1)} \cdots I_{f_n^{-1}(B_n)}$$
$$= I_{f_1^{-1}(B_1) \cap \cdots \cap f_n^{-1}(B_n)}$$
$$= I_{J(f_1,..., f_n)^{-1}(B_1 \times \cdots \times B_n)}$$

We conclude that

$$J(X_{f_1},..., X_{f_n})(B) = I_{J(f_1,..., f_n)^{-1}(B)}$$

for every $B \in \mathscr{B}(\mathbb{R}^n)$. Thus, $J(X_{f_1},..., X_{f_n})$ is the n-dimensional crisp observable corresponding to $J(f_1,..., f_n)$ and we write

$$J(X_{f_1},..., X_{f_n}) = X_{J(f_1,..., f_n)}$$

It follows that the distribution of $J(X_{f_1},..., X_{f_n})$ coincides with the joint distribution of $f_1,..., f_n$.

To summarize, we can describe probability theory in an equivalent way by replacing events by crisp effects ($A \to I_A$), probabilities by expectations ($\mu(A) = \mu(I_A)$), random variables by crisp observables ($f \leftrightarrow X_f$) and Boolean operations by arithmetic operations on crisp effects

$$(I'_A = 1 - I_A, I_A \cap I_B = I_A I_B, I_{A \cup B} = I_A + I_B - I_A I_B)$$

4. FUZZY PROBABILITY THEORY

We now use the ideas of Secs. 2 and 3 to describe fuzzy probability theory. As before, the basic structure is a measurable space $(\Omega, \mathscr{A})$. A random variable $f: \Omega \to [0, 1]$ is called an *effect* or *fuzzy event*. Thus, an effect is just a measurable fuzzy subset of Ω. An effect is *crisp* if it is an indicator function (ordinary probability theory). The set of effects is denoted by $\mathscr{E} = \mathscr{E}(\Omega, \mathscr{A})$. If μ is a probability measure on $(\Omega, \mathscr{A})$ and $f \in \mathscr{E}$, we define the *probability* of f to be its expectation $\mu(f) = \int f\, d\mu$. Notice that $\mu: \mathscr{E} \to \mathbb{R}$ is a probability measure on $\mathscr{E}$ in the following sense. We have $\mu(f) \in [0, 1]$, $\mu(1) = 1$ and if $f \perp g$, then $\mu(f \oplus g) = \mu(f) + \mu(g)$. Also, if $f_i \in \mathscr{E}$ is an increasing sequence, then by the monotone convergence theorem, $\mu(\lim f_i) = \lim \mu(f_i)$ so μ is countably additive. Stated in another way, if a sequence $f_i \in \mathscr{E}$ satisfies $\sum f_i \in \mathscr{E}$, then $\mu(\sum f_i) = \sum \mu(f_i)$.

As in Sec. 2, for $f, g \in \mathscr{E}$, we define $f' = 1 - f$, $f \cap g = fg$ and $f \cup g = f + g - fg$. Notice that f', $f \cap g$ and $f \cup g$ are still elements of $\mathscr{E}$. Applying Lemma 2.4, we obtain the inclusion–exclusion law

$$\mu(f_1 \cup \cdots \cup f_n) = \sum_i \mu(f_i) - \sum_{i<j} \mu(f_i \cap f_j) + \sum_{i<j<k} \mu(f_i \cap f_j \cap f_k) - \cdots + (-1)^{n-1} \mu(f_1 \cap \cdots \cap f_n)$$

As before, we can define a distance $\rho(f, g) = \mu(f \boxplus g)$ and by Corollary 2.9(i), ρ satisfies the triangle inequality $\rho(f, g) \leqslant \rho(f, h) + \rho(h, g)$. However, ρ is not a metric because $\rho(f, g) = 0$ does not imply $f = g$ almost everywhere and if f is not crisp, then $\rho(f, f) \neq 0$.

In Sec. 3 we discussed crisp observables $X_f\colon \mathscr{B}(\mathbb{R}) \to \mathscr{E}_c(\Omega, \mathscr{A})$ and n-dimensional crisp observables $J(X_{f_1},..., X_{f_n})\colon \mathscr{B}(\mathbb{R}^n) \to \mathscr{E}(\Omega, \mathscr{A})$. It is frequently useful to consider more general random variables and crisp observables. Let $(\Lambda, \mathscr{B})$ be another measurable space and let $f\colon \Omega \to \Lambda$ be a measurable function. We call f a *random variable with value space* Λ and the map $X_f\colon \mathscr{B} \to \mathscr{E}_c(\Omega, \mathscr{A})$ given by $X_f(B) = I_{f^{-1}(B)}$ is the corresponding *crisp observable with value space* Λ. We now give the general definition of an observable.

An *observable* or *fuzzy random variable with value* space Λ is a map $X\colon \mathscr{B} \to \mathscr{E}(\Omega, \mathscr{A})$ such that $X(\Lambda) = 1$ and if $B_i \in \mathscr{B}$ are mutually disjoint, then $X(\cup B_i) = \sum X(B_i)$ where the convergence of the summation is pointwise. If $X(B)$ is crisp for every $B \in \mathscr{B}$, then X is *crisp*. If μ is a probability measure on $(\Omega, \mathscr{A})$, then the *distribution* of X is the probability measure μ_X on $(\Lambda, \mathscr{B})$ given by $\mu_X(B) = \mu(X(B))$. Notice that μ_X is indeed a probability measure because $\mu_X(1) = 1$ and if $B_i \in \mathscr{B}$ are mutually disjoint, then by the monotone convergence theorem

$$\mu_X(\cup B_i) = \mu(X(\cup B_i)) = \mu\left(\sum X(B_i)\right) = \sum \mu(X(B_i)) = \sum \mu_X(B_i)$$

For $n \in \mathbb{N}$ we can form the product space $(\Lambda^n, \mathscr{B}^n)$ where $\Lambda^n = \Lambda \times \cdots \times \Lambda$ and $\mathscr{B}^n$ is the σ-algebra on Λ^n generated by the product sets $B_1 \times \cdots \times B_n$. If $X_i\colon \mathscr{B} \to \mathscr{E}(\Omega, \mathscr{A})$, $i = 1,..., n$, are observables, their *joint observable* is the unique observable

$$J(X_1,..., X_n)\colon \mathscr{B}^n \to \mathscr{E}(\Omega, \mathscr{A})$$

with value space Λ^n that satisfies

$$J(X_1,..., X_n)(B_1 \times \cdots \times B_n) = X_1(B_1) \cdots X_n(B_n)$$

for every $B_1, \ldots, B_n \in \mathscr{B}$. Notice that this generalizes our previous definition for crisp observables. The *joint distribution* of $X_1, \ldots, X_n$ is the probability measure $\mu_{X_1, \ldots, X_n}$ on $(\Lambda^n, \mathscr{B}^n)$ given by

$$\mu_{X_1, \ldots, X_n}(B) = \mu_{J(X_1, \ldots, X_n)}(B) = \mu(J(X_1, \ldots, X_n)(B))$$

One can generalize various probabilistic concepts and results concerning random variables to observables. These include independence, conditional expectation, limit laws, convergence theorems and stochastic processes.[4, 6, 10, 13]

There are interesting applications of fuzzy probability theory to quantum mechanics and computer science that we now briefly touch upon Refs. 1, 2, 4, 5, 10, 14. The set of effects $\mathscr{E}(\Omega, \mathscr{A})$ is an example of a σ-effect algebra and these algebras have recently been important in studies of the foundations of quantum mechanics.[3, 7–9, 11] In fact, the term effect was introduced by Ludwig in his work on quantum measurements.[12] We do not need to give the definition of a general σ-effect algebra here because we shall only be concerned with the particular case $\mathscr{E}(\Omega, \mathscr{A})$. An important concept in quantum mechanics is that of a state. In our case, a state on $\mathscr{E}(\Omega, \mathscr{A})$ is a map $s: \mathscr{E}(\Omega, \mathscr{A}) \to [0, 1]$ that satisfies $s(1) = 1$ and if $f_i \in \mathscr{E}(\Omega, \mathscr{A})$ is a sequence such that $\sum f_i \in \mathscr{E}(\Omega, \mathscr{A})$, then $s(\sum f_i) = \sum s(f_i)$. A state s corresponds to a condition or preparation of a system and $s(f)$ is interpreted as the probability that the effect f occurs when the system is in the condition corresponding to s. If μ is a probability measure on $(\Omega, \mathscr{A})$, then it follows from the monotone convergence theorem that $\mu: \mathscr{E}(\Omega, \mathscr{A}) \to [0, 1]$ is a state. In our next result we shall show that every state has this form. Another important concept for σ-effect algebras is that of a σ-morphism. Let $(\Lambda, \mathscr{B})$ be another measurable space. A map $\phi: \mathscr{E}(\Omega, \mathscr{A}) \to \mathscr{E}(\Lambda, \mathscr{B})$ is a *σ-morphism* of $\phi(1) = 1$ and if $f_i \in \mathscr{E}(\Omega, \mathscr{A})$ is a sequence such that $\sum f_i \in \mathscr{E}(\Omega, \mathscr{A})$, then $\phi(\sum f_i) = \sum \phi(f_i)$.

Theorem 4.1. (i) If $\phi: \mathscr{E}(\Omega, \mathscr{A}) \to \mathscr{E}(\Lambda, \mathscr{B})$ is a σ-morphism, then $\phi(\lambda f) = \lambda \phi(f)$ for every $\lambda \in [0, 1]$, $f \in \mathscr{E}(\Omega, \mathscr{A})$. (ii) If $s: \mathscr{E}(\Omega, \mathscr{A}) \to [0, 1]$ is a state, then there exists a unique probability measure μ on $(\Omega, \mathscr{A})$ such that $s(f) = \mu(f)$ for every $f \in \mathscr{E}(\Omega, \mathscr{A})$.

Proof. (i) If $n \in \mathbb{N}$, $f \in \mathscr{E}(\Omega, \mathscr{A})$, then

$$\phi(f) = \phi\left(\frac{1}{n} f + \cdots + \frac{1}{n} f\right) = n\phi\left(\frac{1}{n} f\right) \qquad (n \text{ summands})$$

so $\phi((1/n) f) = (1/n)\, \phi(f)$. If $m, n \in \mathbb{N}$ with $m \leqslant n$, we have

$$\phi\left(\frac{m}{n} f\right) = \phi\left(\frac{1}{n} f + \cdots + \frac{1}{n} f\right) = m\phi\left(\frac{1}{n} f\right) = \frac{m}{n}\, \phi(f) \qquad (m \text{ summands})$$

Hence, $\phi(rf)=r\phi(f)$ for every rational r with $0\leqslant r\leqslant 1$. Let $\lambda\in[0,1]$ be irrational. Then there exists a sequence of rationals $r_i\in[0,1]$ such that $\lambda=\sum r_i$. Since $\sum r_i f=\lambda f\in\mathscr{E}(\Omega,\mathscr{A})$, we have

$$\phi(\lambda f)=\phi\left(\sum r_i f\right)=\sum\phi(r_i f)=\sum r_i\phi(f)=\lambda\phi(f)$$

(ii) Using the same proof as in (i), we have that $s(\lambda f)=\lambda s(f)$ for every $\lambda\in[0,1]$, $f\in\mathscr{E}(\Omega,\mathscr{A})$. Define $\mu\colon\mathscr{A}\to[0,1]$ by $\mu(A)=s(I_A)$. It easily follows that μ is a probability measure. If $f=\sum c_i I_{A_i}$ is a simple function in $\mathscr{E}(\Omega,\mathscr{A})$, we have

$$s(f)=\sum c_i s(I_{A_i})=\sum c_i\mu(A_i)=\mu(f)$$

Since any $f\in\mathscr{E}(\Omega,\mathscr{A})$ is the limit of an increasing sequence of simple functions in $\mathscr{E}(\Omega,\mathscr{A})$, it follows from the countable additivity of s and the monotone convergence theorem that $s(f)=\mu(f)$. For uniqueness, if μ_1 is a probability measure on $(\Omega,\mathscr{A})$ that satisfies $s(f)=\mu_1(f)$ for every $f\in\mathscr{E}(\Omega,\mathscr{A})$, then for every $A\in\mathscr{A}$ we have

$$\mu_1(A)=\mu_1(I_A)=s(I_A)=\mu(I_A)=\mu(A)$$

Hence, $\mu_1=\mu$. □

The next result shows that there exists a natural one-to-one correspondence between observables and σ-morphisms.

Theorem 4.2. If $X\colon\mathscr{B}\to\mathscr{E}(\Omega,\mathscr{A})$ is an observable, then X has a unique extension to a σ-morphism $\tilde{X}\colon\mathscr{E}(\Lambda,\mathscr{B})\to\mathscr{E}(\Omega,\mathscr{A})$. If $Y\colon\mathscr{E}(\Lambda,\mathscr{B})\to\mathscr{E}(\Omega,\mathscr{A})$ is a σ-morphism, then $Y\mid\mathscr{B}$ is an observable.

Proof. Note that $B\mapsto X(B)(\omega)$ is a probability measure on $(\Lambda,\mathscr{B})$ for any $\omega\in\Omega$. For $g\in\mathscr{E}(\Lambda,\mathscr{B})$, define the function $\tilde{X}g$ on Ω by

$$(\tilde{X}g)(\omega)=\int g(\lambda)\,X(d\lambda)(\omega)$$

It is clear that $\tilde{X}$ extends X and that $0\leqslant\tilde{X}g\leqslant 1$. We now show that $\tilde{X}g$ is measurable so that $\tilde{X}\colon\mathscr{E}(\Lambda,\mathscr{B})\to\mathscr{E}(\Omega,\mathscr{A})$. If $g=\sum c_i I_{B_i}$ is a simple function in $\mathscr{E}(\Lambda,\mathscr{B})$, then

$$(\tilde{X}g)(\omega)=\sum c_i X(B_i)(\omega)$$

so $\tilde{X}g$ is measurable. Now for an arbitrary $g \in \mathscr{E}(\Lambda, \mathscr{B})$ there exists an increasing sequence of simple functions $g_i \in \mathscr{E}(\Lambda, \mathscr{B})$ such that $\lim g_i = g$. Then $\tilde{X}g_i$ are measurable, $i = 1, 2, \ldots$, and by the monotone convergence theorem we have

$$(\tilde{X}g)(\omega) = \int \lim g_i(\lambda)\, X(d\lambda)(\omega) = \lim \int g_i(\lambda)\, X(d\lambda)(\omega) = \lim(\tilde{X}g_i)(\omega)$$

Hence, $\tilde{X}g$ is measurable so $\tilde{X}g \in \mathscr{E}(\Omega, \mathscr{A})$. Similar reasoning shows that $\tilde{X}$ is a σ-morphism. For uniqueness, suppose $\phi\colon \mathscr{E}(\Lambda, \mathscr{B}) \to \mathscr{E}(\Omega, \mathscr{A})$ is a σ-morphism that extends X. By Theorem 4.1(i), ϕ and $\tilde{X}$ agree on simple functions and it follows that they coincide on $\mathscr{E}(\Lambda, \mathscr{B})$. The proof of the last statement is straightforward. □

If $f\colon \Omega \to \Lambda$ is a measurable function, the corresponding sharp observable $X_f\colon \mathscr{B} \to \mathscr{E}(\Omega, \mathscr{A})$ is given by $X_f(B) = I_{f^{-1}(B)}$. The next result shows that $X_f\colon \mathscr{E}(\Lambda, \mathscr{B}) \to \mathscr{E}(\Omega, \mathscr{A})$ has a very simple form.

Corollary 4.3. If $f\colon \Omega \to \Lambda$ is a measurable function, then $\tilde{X}_f g = g \circ f$ for every $g \in \mathscr{E}(\Lambda, \mathscr{B})$.

Proof. For $\lambda \in \Lambda$, we denote the Dirac measure concentrated at λ by δ_λ. We then have

$$X_f(B)(\omega) = I_{f^{-1}(B)}(\omega) = \delta_{f(\omega)}(B)$$

Hence, by the proof of Theorem 4.2, for every $g \in \mathscr{E}(\Lambda, \mathscr{B})$ we have

$$(\tilde{X}_f g)(\omega) = \int g(\lambda)\, X_f(d\lambda)(\omega) = \int g(\lambda)\, \delta_{f(\omega)}(d\lambda) = g(f(\omega)) = g \circ f(\omega)$$

It follows that $\tilde{X}_f g = g \circ f$. □

In the sequel, we shall omit the $\sim$ on $\tilde{X}$ and shall frequently identify an observable with its corresponding unique σ-morphism. Let $(\Omega, \mathscr{A})$, $(\Lambda_1, \mathscr{B}_1)$, $(\Lambda_2, \mathscr{B}_2)$ be measurable spaces and let $Y\colon \mathscr{B}_2 \to \mathscr{E}(\Lambda_1, \mathscr{B}_1)$ and $X\colon \mathscr{B}_1 \to \mathscr{E}(\Omega, \mathscr{A})$ be observables. Although we cannot directly compose X and Y, we can compose them if they are thought of as σ-morphisms. Doing this, we have the σ-morphism $X \circ Y\colon \mathscr{E}(\Lambda_2, \mathscr{B}_2) \to \mathscr{E}(\Omega, \mathscr{A})$ which we identify with the observable $X \circ Y\colon \mathscr{B}_2 \to \mathscr{E}(\Omega, \mathscr{A})$. We call $X \circ Y$ the *composition* of X and Y. We then have

$$(X \circ Y)(B)(\omega) = [X(Y(B))](\omega) = \int Y(B)(\lambda_1)\, X(d\lambda_1)(\omega)$$

We close with a discussion of indeterministic automata. In this situation we have an input alphabet Λ_{in}, an output alphabet Λ_{out} and a set of internal states (configurations) Ω of an automaton (computer) M. For simplicity, we assume that Λ_{in}, Λ_{out} and $\Omega=\{\omega_1,\dots,\omega_n\}$ are finite sets. For a deterministic automaton, we have an output function $f_{\text{out}}\colon \Omega\to\Lambda_{\text{out}}$ that prints a symbol in Λ_{out} for every state $\omega\in\Omega$ and a set of functions $\{f_\lambda\colon \lambda\in\Lambda_{\text{in}}\}$ where $f_\lambda\colon \Omega\to\Omega$. If $\lambda\in\Lambda_{\text{in}}$ is input into M and M is in state ω, then M changes to state $f_\lambda(\omega)$. Then $f_{\text{out}}\circ f_\lambda\colon \Omega\to\Lambda_{\text{out}}$ gives the output symbol $f_{\text{out}}\circ f_\lambda(\omega)$ where λ is input and M is in state ω. If we input a program $(\lambda_1,\dots,\lambda_m)$, then the output symbol becomes $f_{\text{out}}\circ f_{\lambda_m}\circ\cdots\circ f_{\lambda_1}(\omega)$.

Suppose now that M is indeterministic. Then with each $\lambda\in\Lambda_{\text{in}}$, ω_j moves to ω_i with probability $g_{\lambda,i}(\omega_j)\in[0,1]$, $i,j=1,\dots,n$. Thus, for every $\lambda\in\Lambda_{\text{in}}$, $g_{\lambda,i}\in\mathscr{E}(\Omega)$ and $\sum_{i=1}^n g_{\lambda,i}=1$. Define the observables $X_\lambda\colon 2^\Omega\to\mathscr{E}(\Omega)$, $\lambda\in\Lambda_{\text{in}}$, by

$$X_\lambda(A)=\sum\{g_{\lambda,i}\colon \omega_i\in A\}$$

Then $X_\lambda(A)$ is the effect that a state moves into A when the input symbol is λ and $X_\lambda(A)(\omega_j)$ is the probability that state ω_j moves into A when the input symbol is λ. Now $f_{\text{out}}^{-1}\colon 2^{\Lambda_{\text{out}}}\to 2^\Omega$ and $X_\lambda\circ f_{\text{out}}^{-1}\colon 2^{\Lambda_{\text{out}}}\to\mathscr{E}(\Omega)$ is the observable given by

$$X_\lambda\circ f_{\text{out}}^{-1}(B)=\sum\{g_{\lambda,i}\colon \omega_i\in f_{\text{out}}^{-1}(B)\}=\sum\{g_{\lambda,i}\colon f_{\text{out}}(\omega_i)\in B\}$$

Thus, $X_\lambda\circ f_{\text{out}}^{-1}(B)$ is the effect that the output is in $B\in 2^{\Lambda_{\text{out}}}$ when the input symbol is λ and $[X_\lambda\circ f_{\text{out}}^{-1}(B)](\omega_j)$ is the probability that a symbol in B is output when λ is input and M is in state ω_j. In particular, for $\alpha\in\Lambda_{\text{out}}$, the probability that α is output when λ is input and M is in state ω_j becomes

$$[X_\lambda\circ f_{\text{out}}^{-1}(\{\alpha\})](\omega_j)=\sum\{g_{\lambda,i}(\omega_j)\colon f_{\text{out}}(\omega_i)=\alpha\}$$

In order to obtain the action of a program $(\lambda_1,\dots,\lambda_m)$ we need the observable $X_{\lambda_1}\circ\cdots\circ X_{\lambda_m}\colon 2^\Omega\to\mathscr{E}(\Omega)$. Then

$$X_{\lambda_1}\circ\cdots\circ X_{\lambda_m}\circ f_{\text{out}}^{-1}\colon 2^{\Lambda_{\text{out}}}\to\mathscr{E}(\Omega)$$

is an observable and $X_{\lambda_1}\circ\cdots\circ X_{\lambda_m}\circ f_{\text{out}}^{-1}(B)$ is the effect that the output is in B when the program $(\lambda_1,\dots,\lambda_m)$ is input. The probabilities can be computed as before. We refer the reader to Refs. 1, 4 for an alternative formulation of these ideas.

We can also consider fuzzy indeterministic automata which do not seem to have been previously discussed. In this case we have a fuzzy output function so we replace f_{out} with an observable $X_{\text{out}}: 2^{\Lambda_{\text{out}}} \to \mathscr{E}(\Omega)$. Then corresponding to a program $(\lambda_1, \ldots, \lambda_m)$ we have an observable

$$X_{\lambda_1} \circ \cdots \circ X_{\lambda_m} \circ X_{\text{out}}: 2^{\Lambda_{\text{out}}} \to \mathscr{E}(\Omega)$$

and the theory proceeds as before.

REFERENCES

1. E. G. Beltrametti and S. Bugajski, "Quantum observables in classical frameworks," *Int. J. Theor. Phys.* **34**, 1221–1229 (1995).
2. E. G. Beltrametti and S. Bugajski, "A classical extension of quantum mechanics," *J. Phys. A: Math. Gen.* **28**, 3329–3334 (1995).
3. E. G. Beltrametti and S. Bugajski, "Effect algebras and statistical physical theories," *J. Math. Phys.* (to appear).
4. S. Bugajski, "Fundamentals of fuzzy probability theory," *Int. J. Theor. Phys.* **35**, 2229–2244 (1996).
5. S. Bugajski, K.-E. Hellwig, and W. Stulpe, "On fuzzy random variables and statistical maps," *Rep. Math. Phys.* (to appear).
6. A. Dvurečenskij and B. Riečan, "On joint distributions of observables for *F*-quantum spaces," *Fuzzy Sets Sys.* **39**, 65–73 (1991).
7. D. Foulis and M. K. Bennett, "Effect algebras and unsharp quantum logics," *Found. Phys.* **24**, 1331–1352 (1994).
8. R. Greechie and D. Foulis, "Transition to effect algebras," *Int. J. Theor. Phys.* **34**, 1369–1382 (1995).
9. S. Gudder, "Examples, problems and results in effect algebras," *Int. J. Theor. Phys.* **35**, 2365–2376 (1996).
10. S. Gudder, "Fuzzy probability theory," *Demon. Math.* **31**, 235–254 (1998).
11. F. Kôpka and F. Chovanec, "*D*-posets," *Math. Slovaca* **44**, 21–34 (1994).
12. G. Ludwig, *Foundations of Quantum Mechanics*, *Vols. I and II* (Springer, Berlin, 1983/1985).
13. R. Mesiar, "Fuzzy observables," *J. Math. Anal. Appl.* **48**, 178–193 (1993).
14. R. Mesiar and K. Piasecki, "Fuzzy observables and fuzzy random variables," *Busefal* **42**, 62–76 (1990).
15. R. Yager, "A note on probabilities of fuzzy events," *Information Sci.* **128**, 113–129 (1979).
16. L. A. Zadeh, "Fuzzy sets," *Information Cont.* **8**, 338–353 (1965).
17. L. A. Zadeh, "Probability measures and fuzzy events," *J. Math. Anal. Appl.* **23**, 421–427 (1968).