9

Generalized Information Theory for Engineering Modeling and Simulation

Cliff Joslyn
Los Alamos National Laboratory

Jane M. Booker
Los Alamos National Laboratory

9.1 Introduction: Uncertainty-Based Information Theory in Modeling and Simulation

Concepts of information have become increasingly important in all branches of science, and especially in modeling and simulation. In the limit, we can view all of science as a kind of modeling. While models can be physical or scale models, more typically we are referring to mathematical or linguistic models, such as $F = ma$, where we measure quantities for mass m and acceleration a, and try to predict another measured quantity force F. More cogently to the readers of this volume, computer simulation models manifest such mathematical formalisms to produce numerical predictions of some technical systems.

A number of points stand out about all models. To quote George Box, "All models are wrong; some models are useful." More particularly:

- All models are necessarily incomplete, in that there are certain aspects of the world which are represented, and others which are not.
- All models are necessarily somewhat in error, in that there will always be some kind of gap between their numerical output and the measured quantities.
- The system being modeled may have inherent variability or un-measurability in its behavior.

0-8493-1180-X/04/$0.00+$1.50

In each case, we wish to be able to measure or quantify these properties, that is, the fidelity and accuracy of our models. We therefore care about the concept of "uncertainty" and all its related concepts: How certain can I be: that I am capturing the properties I'm trying to? that I'm making accurate predictions? that the quantities can be confidently accepted?

We refer to **Uncertainty Quantification** (UQ) as this general task of representing amounts, degrees, and kinds of uncertainty in formal systems. In this context, the concept of uncertainty stands in a dual relation to that of "information." Classically, we understand that when I receive some information, then some question has been answered, and so some uncertainty has been reduced. Thus, this concept of information is that it is a reduction in uncertainty, and we call this **uncertainty-based information**.

Through the 20th century, uncertainty modeling has been dominated by the mathematics of probability, and since Shannon and Weaver [1], information has been defined as a statistical measure of a probability distribution. But also starting in the 1960s, alternative formalisms have arisen. Some of these were intended to stand in contrast to probability theory; others are deeply linked to probability theory but depart from or elaborate on it in various ways. In the intervening time, there has been a proliferation of methodologies, along with concomitant movements to synthesize and generalize them. Together, following Klir [2], we call these **Generalized Information Theory** (GIT).

This chapter surveys some of the most prominent GIT mathematical formalisms in the context of the classical approaches, including probability theory itself. Our emphasis will be primarily on introducing the formal specifications of a range of theories, although we will also take some time to discuss semantics, applications, and implementations.

We begin with the classical approaches, which we can describe as the kinds of mathematics that might be encountered in a typical graduate engineering program. **Logical and set-theoretical** approaches are simply the application of these basic formal descriptions. While we would not normally think of these as a kind of UQ, we will see that in doing so, we gain a great deal of clarity about the other methods to be discussed. We then introduce **interval analysis** and the familiar **probability theory** and related methods.

Following the development of these classical approaches, we move on to consider the GIT proper approaches to UQ. What characterizes a GIT approach is some kind of generalization of or abstraction from a classical approach [3]. **Fuzzy systems theory** was the first and most significant such departure, in which Zadeh generalized the classical, Boolean notions of both set inclusion and truth valuation to representations which are a matter of degree.

A fuzzy set can also be seen as a generalization of a probability distribution or an interval. Similarly, a **monotone** or **fuzzy measure** can be seen as a generalization of a probability measure. A **random set** is a bit different; rather than a generalization, it is an extension of a probability measure to set-valued, rather than point-valued, atomic events. Mathematically, random sets are isomorphic to **Dempster-Shafer bodies of evidence**. Finally, we consider **possibility theory**, which arises as a general alternative to classical information theory based on probability. Possibility measures arise as a different special case of fuzzy measures from probability measures, and are generated in extreme kinds of random sets; similarly, possibility distributions arise as a different special case of fuzzy sets, and generalize classical intervals.

The relations among all the various approaches discussed in this chapter is shown in Figure 9.18. This diagram is somewhat daunting, and so we deliberately show it toward the end of the chapter, after the various subrelations among these components have been explicated. Nonetheless, the intrepid reader might wish to consult this as a reference as the chapter develops.

We also note that our list is not inclusive. Indeed, the field is a dynamic and growing area, with many researchers inventing novel formalisms. Rather, we are trying to capture here the primary classes of GIT theories, albeit necessarily from our perspective. Furthermore, there are a number of significant theoretical components that we will mention in Section 9.4 only in passing, which include:

- **Rough sets** as representations of multi-resolutional structures, and are equivalent to classes of possibility distributions

- Higher-order hybrid structures such as **type II** and **level-II** fuzzy sets and **fuzzified Dempster-Shafer theory**; and finally
- **Choquet capacities** and **imprecise probabilities**, which provide further generalizations of monotone measures.

9.2 Classical Approaches to Information Theory

Throughout this chapter we will assume that we are representing uncertainty claims about some system in the world through reference to a universe of discourse denoted $\Omega = \{\omega\}$. At times we can specify that Ω is finite, countable, or uncountable, depending on the context.

9.2.1 Logical and Set Theoretical Approaches

As mentioned above, some of the most classical mathematical representations can be cast as representations of uncertainty in systems, albeit in a somewhat trivial way. But by beginning this way, we can provide a consistent development of future discussions.

We can begin with a simple proposition A, which may or may not be true of any particular element $\omega \in \Omega$. So if A is true of ω, we can say that the truth value of A for ω is 1: $T_A(\omega) = 1$; and if it is false, that $T_A(\omega) = 0$. Because there are two logical possibilities, 0 and 1, the expression $T_A(\omega)$ expresses the uncertainty, that it might be $T_A(\omega) = 0$, or it might be that $T_A(\omega) = 1$.

Surely the same can be said to be true for any function on Ω. But in this context, it is significant to note the following. First, we *can*, in fact, characterize Boolean logic in this way, characterizing a predicate A as a function $T_A : \Omega \mapsto \{0, 1\}$. The properties of this value set {0, 1} will be crucial below, and will be elaborated on in many of the theories to be introduced.

Second, we can gather together all the $\omega \in \Omega$ for which T_A is true, as distinguished from all those $\omega \in \Omega$ for which T_A is false, and call this the subset $A \subseteq \Omega$, where $A := \{\omega \in \Omega : T_A(\omega) = 1\}$. It is standard to represent the set A in terms of its characteristic function $\chi_A : \Omega \mapsto \{0, 1\}$, where

$$\chi_A(\omega \in \Omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

It is not insignificant that, in fact, $\chi_A \equiv T_A$: the truth value function of the predicate A is equivalent to the characteristic function of the set A. Indeed, there is a mathematical isomorphism between the properties of Boolean logic and those of set theory. For example, the truth table for the logical disjunction ("or") of the two predicates A and B, and the "set disjunction" (union operation) of the two subsets A and B, is shown in Table 9.1. Table 9.2 shows the isomorphic relations among all the primary operations. Graphical representations will be useful below. Letting $\Omega = \{x, y, z, w\}$, Figure 9.1 shows the characteristic function of the subset $A = \{x, z\}$.

TABLE 9.1 Truth Table for Logical Disjunction and Set Disjunction $\cup$

$T_A(\omega)$	$T_B(\omega)$	$T_{A \text{ or } B}(\omega)$	$\chi_A(\omega)$	$\chi_B(\omega)$	$\chi_{A\cup B}(\omega)$
0	0	0	0	0	0
0	1	1	0	1	1
1	0	1	1	0	1
1	1	1	1	1	1

TABLE 9.2 Isomorphisms between Logical and Set Theoretical Operations

Logic		Set Theory	
Negation	$\neg A$	Complement	A^c
Disjunction	A or B	Union	$A \cup B$
Conjunction	A and B	Intersection	$A \cap B$
Implication	$A \to B$	Subset	$A \subseteq B$

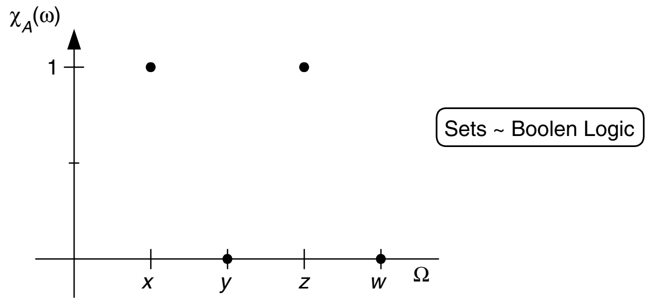

FIGURE 9.1 The crisp set $A = \{x, z\} \subseteq \Omega$.

So far this is quite straightforward, but in so doing we are able to point out the general elements of an uncertainty theory, in particular we can identify:

- The primary objects, in this case sets or propositions, A
- Compound objects as collections of these objects on which the uncertainty can be valued, in this case the power set 2^Ω, that is, the set of all subsets of Ω, so that $2^\Omega = \{A : A \subseteq \Omega\}$
- A range of possible uncertainty quantities for each element $\omega \in \Omega$ with respect to the object $A \in 2^\Omega$, in this case the two binary choices 0 and 1
- Standard operations to combine different objects A and B

The other necessary element is a measure of the total uncertainty or information content $U(A)$ of a particular set or proposition A. For set theory and logic, as well as all the subsequent theories to be presented, such measures are available. Because this is not the primary subject of this chapter, we refer the interested reader elsewhere [4]; nonetheless, for each of the structures we present, we will attempt to identify the community's current best definition for U for that theory. In this case, it is the Hartley measure of information, which is quite simply

$$U_{\text{logic}}(A) = \log_2(|A|), \tag{9.1}$$

where $|\cdot|$ indicates the cardinality of the set. For this and all other uncertainty-based information measures, logarithmic scales are used for their ability to handle addition of two distinct quantities of information, and they are valued in units of bits.

Finally, in each of the cases below, we will identify and observe an **extension principle**, which effectively means that when we generalize from one uncertainty theory to another, then, first, the results from the first must be expressible in terms of special cases of the second, and furthermore the particular properties of the first are recovered exactly for those special cases. However, it is a corollary that in the more general theory, there is typically more than one way to express the concepts that had been previously unequivocal in the more specific theory. This is stated abstractly here, but we will observe a number of particular cases below.

9.2.2 Interval Analysis

As noted, as we move from theory to theory, it may be useful for us to change the properties of the universe of discourse Ω. In particular, in real applications it is common to work with real-valued quantities. Indeed, for many working scientists and engineers, it is always presumed that $\Omega = \mathbb{R}$. The analytical properties of $\mathbb{R}$ are such that further restrictions can be useful. In particular, rather than working with arbitrary subsets of $\mathbb{R}$, it is customary to restrict ourselves to relatively closed sets, specifically closed intervals $I = [I_l, I_u] \subseteq \mathbb{R}$ or half-open intervals $I = [I_l, I_u) \subseteq \mathbb{R}$. Along these lines, it can be valuable to identify

$$\mathcal{D} := \{[a, b) \subseteq \mathcal{R} : a, b \in \mathcal{R}, a \leq b\} \tag{9.2}$$

as the **Borel field** of half-open intervals.

In general, interval-valued quantities represent uncertainty in terms of the upper and lower bounds I_l and I_u. That is, a quantity $x \in \mathbb{R}$ is known to be bounded in this way, such that $x \in [I_l, I_u]$, or $I_l \le x \le I_u$. Because I is a subset of $\mathbb{R}$, it has a characteristic function $\chi_I : \mathbb{R} \mapsto \{0,1\}$, where

$$\chi_I(x) = \begin{cases} 1, & I_l \le x \le I_u \\ 0, & \text{otherwise} \end{cases}$$

The use of intervals generally is well-known in many aspects of computer modeling and simulation [5].[1] We can also observe the components of interval analysis necessary to identify it as an uncertainty theory, in particular:

- The basic objects are the numbers $x \in \mathbb{R}$.
- The compound objects are the intervals $I \subseteq \mathbb{R}$.
- Note that while the universe of discourse has changed from the general set Ω with all its subsets to $\mathbb{R}$ with its intervals, the valuation set has remained $\{0,1\}$.
- The operations are arithmetic manipulations, of the form $I * J$ for two intervals I and J, where $* \in \{+, -, \times, \div, \min, \max\}$, etc. In general, we have

$$I * J := \{x * y : x \in I, y \in J\}. \tag{9.3}$$

  For example, we have

$$I + J = [I_l, I_u] + [J_l, J_u] = \{x + y : x \in I, y \in J\} = [I_l + I_u, J_l + J_u].$$

  Note, however, that in general for an operator $*$ we usually have

$$I * J \ne [I_l * I_u, J_l * J_u].$$

- Finally, for uncertainty we can simply use the width of the interval $U_{\text{int}}(I) := |I| = |I_u - I_l|$, or its logarithm

$$U_{\text{int}}(I) := \log_2(|I|). \tag{9.4}$$

The interval operation [1, 2] + [1.5, 3] = [2.5, 5] is shown in Figure 9.2. Note that

$$U([2.5,\ 5]) = \log_2(2.5) = 1.32 > U([1,\ 2]) + U([1.5,\ 3]) = \log_2(1) + \log_2(1.5) = .58.$$

Also note that interval analysis as such is a kind of set theory: each interval $I \subseteq \mathbb{R}$ is simply a special kind of subset of the special universe of discourse $\mathbb{R}$. Thus we can see in Figure 9.2 that intervals have effectively the same form as the subsets $A \subseteq \Omega$ discussed immediately above in Section 9.2.1, in that they are shown as characteristic functions valued on {0, 1} only.

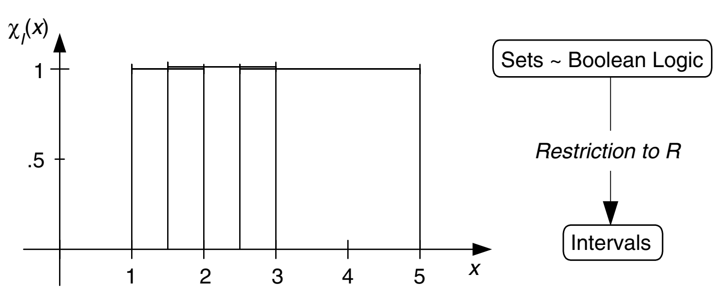

FIGURE 9.2 Interval arithmetic: $[1,2] + [1.5,3] = [2.5,5] \subseteq \mathbb{R}$.

[1]For example, see the journal *Reliable Computing*.

Finally, we can observe the extension principle, in that a number $x \in \mathbb{R}$ can be represented as the degenerate interval $[x, x]$. Then, indeed, we do have that $x * y = [x, x] * [y, y]$.

9.2.3 Probabilistic Representations

By far, the largest and most successful uncertainty theory is probability theory. It has a vast and crucial literature base, and forms the primary point of departure for GIT methods.

Probability concepts date back to the 1500s, to the time of Cardano when gamblers recognized that there were rules of probability in games of chance and, more importantly, that avoiding these rules resulted in a sure loss (i.e., the classic coin toss example of "heads you lose, tails I win," referred to as the "Dutch book"). The concepts were still very much in the limelight in 1685, when the Bishop of Wells wrote a paper that discussed a problem in determining the truth of statements made by two witnesses who were both known to be unreliable to the extent that they only tell the truth with probabilities p_1 and p_2, respectively. The Bishop's answer to this was based on his assumption that the two witnesses were independent sources of information [6].

Mathematical probability theory was initially developed in the 18th century in such landmark treatises as Jacob Bernoulli's *Ars Conjectandi* (1713) and Abraham DeMoiver's *Doctrine of Chances* (1718, 2nd edition 1738). Later in that century, articles would appear that provided the foundations of modern interpretations of probability: Thomas Bayes' "An Essay Towards Solving a Problem in the Doctrine of Chances," published in 1763 [7], and Pierre Simon Laplace's formulation of the axioms relating to games of chance, "Memoire sur la Probabilite des Causes par les Evenemens," published in 1774. In 1772, the youthful Laplace began his work in mathematical statistics and provided the roots for modern decision theory.

By the time of Newton, physicists and mathematicians were formulating different theories of probability. The most popular ones remaining today are the relative frequency theory and the subjectivist or personalistic theory. The latter development was initiated by Thomas Bayes [7], who articulated his very powerful theorem, paving the way for the assessment of subjective probabilities. The theorem gave birth to a subjective interpretation of probability theory, through which a human's degree of belief could be subjected to a coherent and measurable mathematical framework within the subjective probability theory.

9.2.3.1 Probability Theory as a Kind of GIT

The mathematical basis of probability theory is well known. Its basics were well established by the early 20th century, when Rescher developed a formal framework for a conditional probability theory and Jan Lukasiewicz developed a multivalued, discrete logic circa 1930. But it took Kolmogorov in the 1950s to provide a truly sound mathematical basis in terms of measure theory [8].

Here we focus only on the basics, with special attention given to casting probability theory in the context of our general development of GIT.

In our discussions on logic and set theory, we relied on a general universe of discourse Ω and all its subsets $A \in 2^\Omega$, and in interval analysis we used $\mathbb{R}$ and the set of all intervals $I \in \mathcal{D}$. In probability theory, we return to a general universe of discourse Ω, but then define a **Boolean field** $\mathcal{E} \subseteq 2^\Omega$ on Ω as a collection of subsets closed under union and intersection:

$$\forall A_1, A_2 \in \mathcal{E}, \quad A_1 \cup A_2 \in \mathcal{E}, \quad A_1 \cap A_2 \in \mathcal{E}.$$

Note that while 2^Ω is a field, there are fields which are not 2^Ω.

We then define a probability measure Pr on $\mathcal{E}$ as a set function $\Pr : \mathcal{E} \mapsto [0, 1]$ where $\Pr(\Omega) = 1$ as a **normalization** condition, and if $A_1, A_2, \ldots$ is a countably infinite sequence of mutually disjoint sets in $\mathcal{E}$ whose union is in $\mathcal{E}$, then

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i).$$

When all of these components are in place, we call the collection $\langle \Omega, \mathcal{E}, \text{Pr} \rangle$ a **probability space**, where we use $\langle \cdot \rangle$ to indicate a general *n*-tuple, in this case an ordered triple.

We interpret the probability of an event $A \in \mathcal{E}$ as the uncertainty associated with the outcome of A considered as an event. Implicit in this definition are two additional concepts, time, t, and "history" or background information, H, available for contemplating the uncertain events, at t and H. Thus we can also use the revised notation $\text{Pr}(A;\, H,\, t)$. Below, we will freely use either notation, depending on the context.

The calculus of probability consists of certain rules (or axioms) denoted by a number determined by $\text{Pr}(A; H, t)$, in which the probability of an event, A, is related to H at time t. When the event A pertains to the ability to perform a certain function (e.g., survive a specified mission time), then $\text{Pr}(A; H, t)$ is known as the product's reliability. This is a traditional definition of reliability, although we must note that treatments outside of the context of probability theory, indeed, outside of the context of any uncertainty-based information theory, are also possible [9].

The quantity $\text{Pr}(A_1 \,|\, A_2; H, t)$ is known as the conditional probability of A_1, given A_2. Note that conditional probabilities are in the subjunctive. In other words, the disposition of A_2 at time t, were it to be known, would become a part of the history H at time t. The vertical line between A_1 and A_2 represents a supposition or assumption about the occurrence of A_2.

We can also define a function called a **probability distribution** or **density**, depending on the context, as the probability measure at a particular point $\omega \in \Omega$. Specifically, we have $p\colon\ \Omega \mapsto [0,1]$ where $\forall \omega \in \Omega$, $p(\omega) := \text{Pr}(\{\omega\})$. When Ω is finite, then we tend to call p a discrete distribution, and we have

$$\forall A \subseteq \Omega, \quad \text{Pr}(A) = \sum_{\omega \in A} p(\omega),$$

and as the normalization property we have $\sum_{\omega \in \Omega} p(\omega) = 1$.

When $\Omega = \mathbb{R}$, then we tend to use f, and call it a **probability density function** (pdf). We then have

$$\forall A \subseteq \mathbb{R}, \quad \text{Pr}(A) = \int_A f(x)dx,$$

and for normalization $\int_{-\infty}^{\infty} f(x)dx = 1$. In this case, we can also define the **cumulative distribution** as

$$\forall x \in \mathbb{R}, \quad F(x) := \text{Pr}\,((-\infty, x]) = \int_{-\infty}^{x} f(x)dx,$$

and for normalization $\lim_{x \to \infty} F(x) = 1$.

To make the terminological problems worse, it is common to refer to the cumulative distribution as simply the "distribution function." These terms, especially "distribution," appear frequently below in different contexts, and we will try to use them clearly.

We have now introduced the basic components of probability theory as a GIT:

- The objects are the points $\omega \in \Omega$.
- The compound objects are the sets in the field $A \in \mathcal{E}$.
- The valuation set has become the unit interval [0, 1].

We are now prepared to introduce the operations on these objects, similar to logic and intervals above. First we exploit the isomorphism between sets and logic by introducing the formulation

$$\text{Pr}(A \text{ or } B; H,\ t) := \text{Pr}(A \cup B; H,\ t), \quad \text{Pr}(A \text{ and } B; H,\ t) := \text{Pr}(A \cap B; H,\ t).$$

The calculus of probability consists of the following three primary rules:

1. **Convexity:** For any event $A \in \mathcal{E}$, we have $0 \leq \Pr(A; H, t) \leq 1$. Note that this is effectively a restatement of the definition, since $\Pr: 2^{\Omega} \mapsto [0, 1]$.
2. **Addition:** Assume two events A_1 and A_2 that are mutually exclusive; that is, they cannot simultaneously take place, so that $A_1 \cap A_2 = \varnothing$. Then we have

$$\Pr(A_1 \text{ or } A_2; H, t) = \Pr(A_1; H, t) + \Pr(A_2; H, t).$$

   In general, for any two sets $A, B \in \mathcal{E}$, we have

$$\Pr(A_1 \text{ or } A_2; H, t) = \Pr(A_1; H, t) + \Pr(A_2; H, t) - \Pr(A_1 \cap A_2; H, t). \tag{9.5}$$

3. **Multiplication:** Interpreting $\Pr(A_1 \mid A_2; H, t)$ as a quantification of the uncertainty about an event A_1 supposing that event A_2 has taken place, then we have

$$\Pr(A_1 \text{ and } A_2; H, t) = \Pr(A_1 \mid A_2; H, t) \Pr(A_2; H, t).$$

Finally, $\Pr(A_1 \text{ and } A_2; H, t)$ also can be written as $\Pr(A_2 \mid A_1; H, t) \Pr(A_1; H, t)$ because at time t both A_1 and A_2 are uncertain events and one can contemplate the uncertainty about A_1 supposing that A_2 were to be true or vice versa.

To complete the characterization of probability theory as a GIT, we can define the total uncertainty for a discrete as its statistical entropy:

$$U_{\text{prob}}(p) = -\sum_{\omega \in \Omega} p(\omega) \log_2(p(\omega)). \tag{9.6}$$

It is not so straightforward for a continuous pdf, but these concepts are related to variance and other measures of the "spread" or "width" of the density.

9.2.3.2 Interpretations of Probability

The calculus of probability does not tell us how to interpret probability, nor does the theory define what probability means. The theory and calculus simply provide a set of rules by which the uncertainties about two or more events combine or "cohere." Any set of rules for combining uncertainties that are in violation of the rules given above are said to be "incoherent" or inconsistent with respect to the calculus of probability. But it is crucial to note that it is exactly these "inconsistencies" that have spurred much of the work in generalizing the structures reported herein.

Historically speaking, there have been at least 11 different significant interpretations of probability; the most common today are relative frequency theory and personalistic or subjective theory.

Relative frequency theory has its origins dating back to Aristotle, Venn, von Mises, and Reichenbach. In this interpretation, probability is a measure of an empirical, objective, and physical fact of the world, independent of human knowledge, models, and simulations. Von Mises believed probability to be a part of a descriptive model, whereas Reichenbach viewed it as part of the theoretical structure of physics. Because probability is based only on observations, it can be known only *a posteriori* (literally, after observation). The core of this interpretation is in the concept of a random collective, as in the probability of finding an ace in a deck of cards (the collective). In relative frequency theory, $\Pr(A; H, t) = \Pr(A)$; there is no H or t.

Personalistic or subjective interpretation of probability has its origins attributed to Borel, Ramsey, de Finetti, and Savage. According to this interpretation, there is no such thing as a correct probability, an unknown probability, or an objective probability. Probability is defined as a degree of belief, or a willingness to bet: the probability of an event is the amount (say p) the individual is willing to bet, on a two-sided bet, in exchange for $1, should the event take place. By a two-sided bet is meant staking

$(1-p)$ in exchange for \$1, should the event not take place. Probabilities of one-of-a-kind or rare events, such as the probability of intelligent life on other planets, are easily handled with this interpretation.

The personalistic or subjective probability permits the use of all forms of data, knowledge, and information. Therefore, its usefulness in applications where the required relative frequency data are absent or sparse becomes clear. This view of probability also includes Bayes theorem and comes the closest of all the views of probability to the interpretation traditionally used in fuzzy logic. Therefore, this interpretation of probability can be the most appropriate for addressing the uncertainties in complex decisions surrounding modern reliability problems.

9.2.3.3 Bayes Theorem and Likelihood Approaches for Probability

In 1763, the Reverend Thomas Bayes of England made a momentous contribution to probability, describing a relationship among probabilities of events (A_1 and A_2) in terms of conditional probability:

$$\Pr(A_1 \mid A_2;\, H) = \frac{\Pr(A_2 \mid A_1;\, H)\Pr(A_1;\, H)}{\Pr(A_2;\, H)}.$$

Its development stems from the third "multiplication" axiom of probability defining conditional probability. Bayes theorem expresses the probability that event A_1 occurs if we have observed A_2 in terms of the probability of A_2 given that A_1 occurred.

Historical investigation reveals that Laplace may have independently established another form of Bayes Theorem by considering A_1 as a comprised of k sub-events, $A_{11}, A_{12}, \ldots, A_{1k}$. Then the probability of $A_2, \Pr(A_2;\, H)$ can be rewritten as

$$\Pr(A_2 \mid A_{11};\, H)\Pr(A_{11};\, H) + \Pr(A_2 \mid A_{12};\, H)\Pr(A_{12};\, H) + \cdots + \Pr(A_2 \mid A_{1k};\, H)\Pr(A_{1k};\, H).$$

This relationship is known as the **Law of Total Probability** for two events, A_1 and A_2, and can be rewritten as:

$$\Pr(A_2 = a_2;\, H) = \sum_j \Pr(A_2 = a_2 \mid A_1 = a_{1j};\, H)\Pr(A_1 = a_{1j};\, H),$$

where lower case a values are particular values or realizations of the two events.

The implications of Bayes theorem are considerable in its use, flexibility, and interpretation in that [10]:

- It demonstrates the proportional relationship between the conditional probability $\Pr(A_1 \mid A_2;\, H)$ and the product of probabilities $\Pr(A_1;\, H)$ and $\Pr(A_2 \mid A_1;\, H)$.
- It prescribes how to relate the two uncertainties about A_1: one prior to knowing A_2, the other posterior to knowing A_2.
- It specifies how to change the opinion about A_1 were A_2 to be known; this is also called "the mathematics of changing your mind".
- It provides a mathematical way to incorporate additional information.
- It defines a procedure for the assessor, i.e., how to bet on A_1 should A_2 be observed or known. That is, it prescribes the assessor's behavior before actually observing A_2.

Because of these implications, the use of Bayesian methods, from the application of this powerful theorem, have become widespread as an information combination scheme and as an updating tool, combining or updating the prior information with the existing information about events. These methods also provide a mechanism for handling different kinds of uncertainties within a complex problem by linking subjective-based probability theory and fuzzy logic.

The prior about A_1 refers to the knowledge that exists prior to acquisition of information about event A_1. The fundamental Bayesian philosophy is that prior information is valuable, should be used, and can be mathematically combined with new or updating information. With this combination, uncertainties can be reduced.

Bernoulli appears to be the first to prescribe uncertainty about A_1 if one were to observe A_2 (but assuming $A_2 = a_2$ has not yet occurred). By dropping the denominator and noting the proportionality of the remaining terms on the right-hand side, Bayes rule becomes:

$$\Pr(A_1 \mid A_2; H) \propto \Pr(A_2 \mid A_1; H)\Pr(A_1; H).$$

However, $A_2 = a_2$ is actually observed, making the left-hand side written as $\Pr(A_1; a_2, H)$. Therefore,

$$\Pr(A_1;\ a_2,\ H) \propto \Pr(A_2 = a_2 \mid A_1 = a_1;\ H)\Pr(A_1 = a_1;\ H).$$

However, there is a problem because $\Pr(A_2 = a_2 \mid A_1 = a_1; H)$ is no longer interpreted as a probability. Instead, this term is called the **likelihood** that $A_1 = a_1$ in light of H and the fact that $A_2 = a_2$. This is denoted $L(A_1 = a_1; a_2, H)$. This likelihood is a function of a_1 for a fixed value of a_2. For example, the likelihood of a test resulting in a particular failure rate would be expressed in terms of $L(A_1 = a_1; a_2, H)$.

The concept of a likelihood gives rise to another formulation of Bayes theorem:

$$\Pr(A_1; a_2, H) \propto L(A_1 = a_1; a_2, H)\Pr(A_1 = a_1; H).$$

Here, $\Pr(A_1 = a_1; H)$ is again the *prior probability* of A_1 (i.e., the source for information that exists "prior" to test data (a_2) in the form of expert judgment and other historical information). By definition, the prior represents the possible values and associated probabilities for the quantity of interest, A_1. For example, one decision is to represent the average failure rate of a particular manufactured item. The likelihood $L(A_1 = a_1; a_2, H)$ is formed from data in testing a specified number of items. Test data from a previously made item similar in design forms the prior. $\Pr(A_1; a_2, H)$ is the posterior distribution in the light of a_2 (the data) and H, produced from the prior information and the data.

The likelihood is an intriguing concept but it is not a probability, and therefore does not obey the axioms or calculus of probability. In Bayes theorem, the likelihood is a connecting mechanism between the two probabilities: the prior probability, $\Pr(A_1; H)$, and the posterior probability, $\Pr(A_1; a_2, H)$. The likelihood is a subjective construct that enables the assignment of relative weights to different values of $A_1 = a_1$.

9.2.3.4 Distribution Function Formulation of Bayes Theorem

Bayes theorem has been provided for the discrete form for two random variables representing the uncertain outcomes of two events, A_1 and A_2. For continuous variables X and Y, the probability statements are replaced by pdfs, and the likelihood is replaced by a likelihood function. If Y is a continuous random variable whose probability density function depends on the variable X, then the conditional pdf of Y given X is $f(y \mid x)$. If the prior pdf of X is $g(x)$, then for every y such that $f(y) > 0$ exists, the posterior pdf of X, given $Y = y$ is

$$g(x \mid y; H) = \frac{f(y \mid x;\ H)g(x;\ H)}{\int f(y \mid x;\ H)g(x;\ H)dx},$$

where the denominator integral is a normalizing factor so that $g(x \mid y; H)$, the posterior distribution, integrates to 1 (as a proper pdf).

Alternatively, utilizing the likelihood notation, we have

$$g(x \mid y; H) \propto L(x \mid y; H)g(x; H),$$

so that the posterior is proportional to the likelihood function times the prior distribution.

In this form, Bayes theorem can be interpreted as a weighting mechanism. The theorem mathematically weights the likelihood function and prior distribution, combining them to form the posterior. If these two distributions overlap to a large extent, this mathematical combination produces a desirable result: the uncertainty (specifically, the variance) of the posterior distribution is smaller than that produced by a simple weighted combination, $w_1 g_{\text{prior}} + w_2 L_{\text{likelihood}}$, for example. The reduction in the uncertainty results from the added information of combining two distributions that contain similar information (overlap).

Contrarily, if the prior and likelihood are widely separated, then the posterior will fall in the gap between the two functions. This is an undesirable outcome because the resulting combination falls in a region unsupported by either the prior or the likelihood. In this situation, one would want to reconsider using Bayesian combination and either seek to resolve the differences between the prior and likelihood or use some other combination method such as a simple weighting scheme; for example, consider $w_1 g_{\text{prior}} + w_2 L_{\text{likelihood}}$.

As noted above, a major advantage of using Bayes theorem to combine distribution functions of different information sources is that the spread (uncertainty) in the posterior distribution is reduced when the information in the prior and likelihood distributions are consistent with each other. That is, the combined information from the prior distribution and the data has less uncertainty because the prior distribution and data are two different information sources that support each other.

Before the days of modern computers and software, calculating Bayes theorem was computationally cumbersome. For those times, it was fortunate that certain choices of pdfs for the prior and likelihood produced easily obtained posterior distributions. For example, a beta prior with a binomial likelihood produces a beta posterior whose parameters are simple functions of the prior beta and binomial parameters, as the following example illustrates. With modern computational methods, these analytical shortcuts, called conjugate priors, are not necessary; however, computation still has its difficulties in how to formulate, sample from, and parameterize the various functions in the theorem. Simulation algorithms such as Metropolis-Hastings and Gibbs sampling provide the how-to, but numerical instabilities and convergence problems can occur with their use. A popular simulation technique for simulation and sampling is Markov Chain Monte Carlo (MCMC). A flexible software package for implementation written in Java is YADAS[2] [11].

9.2.3.5 Binomial/Beta Reliability Example

Suppose we prototype a system, building 20 units, and subject these to a stress test. All 20 units pass the test [10]. The estimate of success/failure rates from test data alone is $n_1 = 20$ tests with $x_1 = 20$ successes.

Using just this information, the success rate is 20/20 = 1, and the failure rate is 0/20 = 0. This fundamental reliability (frequentist interpretation of probability) estimate, based on only 20 units, does not reflect the uncertainty in the reliability for the system and does not account for any previously existing information about the units before the test.

A Bayesian approach can take advantage of prior information and provide an uncertainty estimate on the probability of a success, p. Prior knowledge could exist in many forms: expertise of the designers, relevant data from similar systems or components, design specifications, historical experience with similar designs, etc. that can be used to formulate the prior distribution for p, $g(p)$. The beta distribution is often chosen as a prior for a probability because it ranges from 0 to 1 and can take on many shapes (uniform, "J" shape, "U" shape and Gaussian-like) by adjusting its two parameters, n_0 and x_0. That beta prior is denoted as beta(x_0, n_0), and its pdf is:

$$g(p) = \frac{\Gamma(n_0)}{\Gamma(x_0)\Gamma(n_0 - x_0)} x^{x_0 - 1}(1-x)^{n_0 - x_0 - 1}.$$

[2]Yet Another Data Analysis System.

For this example, assume the prior information is in the form of an estimate of the failure rate from the test data done on a similar system that is considered relevant for this new system with $n_0 = 48$ tests on a similar system with $x_0 = 47$ successes.

The new prototype test data forms the likelihood, $L(p; x)$. Because this data represents the number of successes, x_1, in n_1 trials it conforms to the binomial distribution with the parameter of interest for success, p. The beta distribution, $g(p)$, is a conjugate prior when combined with the binomial likelihood, $L(p; x)$, using Bayes theorem. Thus, the resulting, posterior distribution, $g(p \mid x)$ is also a beta distribution with parameters $(x_0 + x_1, n_0 + n_1)$:

$$g(p \mid x) = \frac{\Gamma(n_0 + n_1)}{\Gamma(x_0 + x_1)\Gamma(n_0 + n_1 - x_0 - x_1)} p^{x_0 + x_1 - 1}(1 - p)^{n_0 + n_1 - x_0 - x_1 - 1},$$

or

$$g(p \mid x) = \frac{\Gamma(68)}{\Gamma(67)\Gamma(1)} p^{66}(1 - p)^{0}.$$

The mean success rate of the beta posterior is

$$\frac{x_0 + x_1}{n_0 + n_1} = \frac{67}{68} = 0.985,$$

or, in terms of a mean failure rate for the beta posterior, approximately $1 - 0.985 = 0.015$ failure rate. The variance of the beta posterior distribution is:

$$\frac{(x_0 + x_1)[(n_0 + n_1) - (x_0 + x_1)]}{(n_0 + n_1)^2 (n_0 + n_1 + 1)} = 0.00021.$$

The engineering reliability community gravitates to the binomial/beta conjugate prior because many of the failures are binomial in nature and the parameters of the prior and posterior can have a reliability-based interpretation: n_0 = number of tests and x_0 = number of successes for the prior parameter interpretation. Similarly, $n_0 + n_1$ = number of pseudo tests and $x_0 + x_1$ = number of pseudo successes for the posterior parameter interpretation, provided these values are greater than 1.

9.3 Generalized Information Theory

We now turn our attention to the sub-fields of GIT proper. Most of these formalisms were developed in the context of probability theory, and are departures, in the sense of generalization from or elaborations, of it. However, many of them are also intricately interlinked with logic, set theory, interval analysis, combinations of these, combinations with probability theory, and combinations with each other. We emphasize again that there is a vast literature on these subjects in general, and different researchers have different views on which theories are the most significant, and how they are related. Our task here is to represent the primary GIT fields and their relations in the context of probability theory and reliability analysis. For more background, see work elsewhere [12–15].

9.3.1 Historical Development of GIT

As mentioned in the introduction, we will describe the GIT sub-fields of fuzzy systems, monotone or fuzzy measures, random sets, and possibility theory. While these GIT sub-fields developed historically in the context of probability theory, each also has progenitors in other parts of mathematics.

In 1965, Lotfi Zadeh introduced his seminal idea in a continuous-valued logic called fuzzy set theory [16, 17]. In doing so, he was recapitulating some earlier ideas in multi-valued logics [13].

Also in the 1960s, Arthur Dempster developed a statistical theory of evidence based on probability distributions propagated through multi-valued maps [18]. In so doing, he introduced mathematical structures that had been identified by Choquet some years earlier, and described as general "capacities." These Choquet capacities [19, 20] are generalizations of probability measures, as we shall describe below. In the 1970s, Glenn Shafer extended Dempster's work to produce a complete theory of evidence dealing with information from more than one source [21]. Since then, the combined sub-field has come to be known as "Dempster-Shafer Evidence Theory" (DS Theory). Meanwhile, the stochastic geometry community was exploring the properties of random variables valued not in $\mathbb{R}$, but in closed, bounded subsets of $\mathbb{R}^n$. The random sets they described [22, 23] turned out to be mathematically isomorphic to DS structures, although again, with somewhat different semantics. This hybrid sub-field involving Dempster's and Shafer's theories and random sets all exist in the context of infinite-order Choquet capacities.

While random sets are defined in general on $\mathbb{R}^n$, for practical purposes, as we have seen, it can be useful to restrict ourselves to closed or half-open intervals of $\mathbb{R}$ and similar structures for $\mathbb{R}^n$. Such structures provide DS correlates to structures familiar to us from probability theory as it is used, for example pdfs and cumulative distributions. It should be noted that Dempster had previously introduced such "random intervals" [24].

In 1972, Sugeno introduced the idea of a "fuzzy measure" [25], which was intended as a direct generalization of probability measures to relax the additivity requirement of additive Equation 9.5. Various classes of fuzzy measures were identified, many of the most useful of which were already available within DS theory. In 1978, Zadeh introduced the special class of fuzzy measures called "possibility measures," and furthermore suggested a close connection to fuzzy sets [26]. It should be noted that there are other interpretations of this relation, and to some extent it is a bit of a terminological oddity that the term "fuzzy" is used in two such contexts [27]. For this reason, researchers are coming to identify fuzzy measures instead as "monotone measures."

The 1980s and 1990s were marked by a period of synthesis and consolidation, as researchers completed some open questions and continued both to explore novel formalisms, but more importantly the relations among these various formalisms. For example, investigators showed a strong relationship between evidence theory, probability theory, and possibility theory with fuzzy measures [28].

One of the most significant developments during this period was the introduction by Walley of an even broader mathematical theory of "imprecise probabilities," which further generalizes fuzzy measures [29].

9.3.2 GIT Operators

The departure of the GIT method from probability theory is most obviously significant in its use of a broader class of mathematical operators. In particular, probability theory operates through the familiar algebraic operators addition + and multiplication ×, as manifested in standard linear algebra. GIT recognizes + as an example of a generalized disjunction, the "or"-type operator, and × as an example of a generalized conjunction, the "and"-type operator, but uses other such operators as well.

In particular, we can define the following operations:

Complement: Let c: $[0, 1] \mapsto [0, 1]$ be a **complement** function when

$$c(0) = 1, \quad c(1) = 0, \quad x \le y \mapsto c(x) \ge c(y).$$

Norms and Conorms: Assume associative, commutative functions $\sqcap : [0,1]^2 \mapsto [0,1]$ and $\sqcup : [0,1]^2 \mapsto [0,1]$. Because of associativity, we can use the operator notation $x \sqcap y := \sqcap(x, y)$, $x \sqcup y := \sqcup(x, y)$. Further assume that $\sqcup$ and $\sqcap$ are monotonic, in that

$$\forall x \le y, z \le w, \quad x \sqcap z \le y \sqcap w, \quad x \sqcup z \le y \sqcap w.$$

Then $\sqcap$ is a **triangular norm** if it has identity 1, with $1 \sqcap x = x \sqcap 1 = x$; and $\sqcup$ is a **triangular conorm** if it has identity 0, with $0 \sqcup x = x \sqcup 0 = x$.

While there are others, the prototypical complement function, and by far the most commonly used, is $c(x) = 1 - x$. Semantically, complement functions are used for logical negation and set complementation.

In general, there are many continuously parameterized classes of norms and conorms [13]. However, we can identify some typical norms and conorms that may be familiar to us from other contexts. Below, use $\wedge, \vee$ for the maximum and minimum operators, let $x, y \in [0, 1]$, and let $\lfloor x \rfloor$ be the greatest integer below $x \in \mathbb{R}$, and similarly $\lceil x \rceil$ the least integer above x. Then we have:

Norms:
- **–Min:** $x \wedge y$
- **–Times:** $x \times y$
- **–Bounded Difference:** $x -_b y := (x + y - 1) \vee 0$
- **–Extreme Norm:** $\lfloor x \rfloor \times \lfloor y \rfloor$

Conorms:
- **–Max:** $x \vee y$
- **–Probabilistic Sum:** $x +_p y := x + y - xy$
- **–Bounded Sum:** $x +_b y := (x + y) \wedge 1$
- **–Extreme Conorm:** $\lfloor x \rfloor \times \lfloor y \rfloor$

In general, $\wedge$ is the greatest and $\lfloor x \rfloor \times \lfloor y \rfloor$ the least norm, and $\vee$ is the least and $\lceil x \rceil \times \lceil y \rceil$ the greatest conorm. The relations are summarized in Table 9.3.

9.3.3 Fuzzy Systems

In 1965, Zadeh published a new set theory that addressed the kind of vague uncertainty that can be associated with classifying an event into a set [16, 17]. The idea suggested that *set membership* is the key to decision making when faced with linguistic and nonrandom uncertainty. Unlike probability theory, based upon crisp sets, which demands that any outcome of an event or experiment belongs to a set A or to its complement, A^c, and not both, fuzzy set theory permits such a joint membership. The degree of membership that an item belongs to any set is specified using a construct of fuzzy set theory, a membership function.

Just as we have seen that classical (crisp) sets are isomorphic to classical (crisp) logic, so there is a fuzzy logic that is isomorphic to fuzzy sets. Together, we can thus describe **fuzzy systems** as systems whose operations and logic are governed by these principles. Indeed, it can be more accurate to think of a *process* of "fuzzification," in which a formalism that has crisp, binary, or Boolean choices are relaxed to admit degrees of gradation. In this way, we can conceive of such ideas as fuzzified arithmetic, fuzzified calculus, etc.

9.3.3.1 Fuzzy Sets

Zadeh's fundamental insight was to relax the definition of set membership. Where crisp sets contain objects that satisfy precise properties of membership, fuzzy sets contain objects that satisfy imprecise properties of membership; that is, membership of an object in a fuzzy set can be approximate or partial.

We now introduce the basic formalism of fuzzy sets. First, we work with a general universe of discourse Ω. We then define a **membership function** very simply as any function $\mu \colon \Omega \mapsto [0, 1]$. Note, in particular, that the characteristic function χ_A of a subset $A \subseteq \Omega$ is a membership function, simply because $\{0, 1\} \subseteq [0, 1]$.

TABLE 9.3 Prototypical Norms and Conorms

Triangular Norm	$x \sqcap y$:	$x \wedge y$	$\geq$	$x \times y$	$\geq$	$0 \vee (x + y - 1)$	$\geq$	$\lfloor x \rfloor \times \lfloor y \rfloor$
Triangular Conorm	$x \sqcup y$:	$x \vee y$	$\leq$	$x + y - xy$	$\leq$	$1 \wedge (x + y)$	$\leq$	$\lceil x \rceil \times \lceil y \rceil$

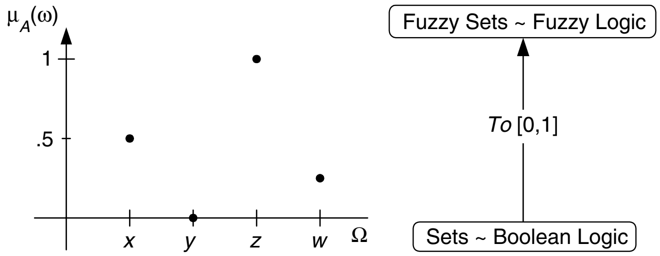

FIGURE 9.3 The fuzzy set $\tilde{A} = \{\langle x, .5\rangle, \langle y, 0\rangle, \langle z, 1\rangle, \langle w, .25\rangle\} \tilde{\subseteq} \Omega$.

Thus, membership functions generalize characteristic functions, and in this way, we can conceive of a **fuzzy subset** of Ω, denoted $\tilde{A} \tilde{\subseteq} \Omega$, as being defined by some particular membership function $\mu_{\tilde{A}}$. For a characteristic function χ_A of a subset A, we interpret $\chi_A(\omega)$ as being 1 if $\omega \in A$, and 0 if $\omega \notin A$. For a membership function $\mu_{\tilde{A}}$ of a fuzzy subset $\tilde{A}$, it is thereby natural to interpret $\mu_{\tilde{A}}(\omega)$ as the *degree* or *extent to which* $\omega \in \tilde{A}$.

Then note especially how the extension principle holds. In particular, consider $\omega_1, \omega_2 \in \Omega$ such that $\mu_{\tilde{A}}(\omega_1) = 0$ and $\mu_{\tilde{A}}(\omega_2) = 1$. In these cases, we can still consider that $\omega_1 \in \tilde{A}$ and $\omega_2 \notin \tilde{A}$ unequivocally.

Below, it will frequently be convenient to denote $\tilde{A}(\omega)$ for $\mu_{\tilde{A}}(\omega)$. Moreover, when $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is finite, we can denote a fuzzy subset as a set of ordered pairs:

$$\tilde{A} = \{\langle \omega_1, \tilde{A}(\omega_1)\rangle, \langle \omega_2, \tilde{A}(\omega_2)\rangle, \ldots, \langle \omega_n, \tilde{A}(\omega_n)\rangle\}.$$

Consider the simple example shown in Figure 9.3. For $\Omega = \{x, y, z, w\}$, we might have

$$\tilde{A} = \{\langle x, .5\rangle, \langle y, 0\rangle, \langle z, 1\rangle, \langle w, .25\rangle\},$$

so that z is completely in $\tilde{A}$, y is completely not in $\tilde{A}$, and x and w are in $\tilde{A}$ to the intermediate extents .5 and .25, respectively.

When a membership function actually reaches the line $\mu = 1$, so that $\exists \omega \in \Omega, \tilde{A}(\omega) = 1$, then we call $\tilde{A}$ **normal**. This usage is a bit unfortunate because it may indicate probabilistic additive normalization, so we will try to distinguish this as fuzzy normalization. Fuzzy normalization is also the criterion for $\tilde{A}$ to be a possibility distribution, which we will discuss below in Section 9.3.6.

To continue our characterization of fuzzy sets as a GIT, we need to define the correlates to the basic set operations. Not surprisingly, we will do this through the generalized operators introduced in Section 9.3.2. Below, presume two fuzzy sets $\tilde{A}, \tilde{B} \tilde{\subseteq} \Omega$. Then we have:

Fuzzy Complement: $\mu_{\tilde{A}^c}(\omega) = c(\tilde{A}(\omega))$
Fuzzy Union: $\mu_{\tilde{A} \cup \tilde{B}}(\omega) = \tilde{A}(\omega) \sqcup \tilde{B}(\omega)$
Fuzzy Intersection: $\mu_{\tilde{A} \cap \tilde{B}}(\omega) = \tilde{A}(\omega) \sqcap \tilde{B}(\omega)$
Fuzzy Set Equivalence: $\tilde{A} = \tilde{B} := \forall \omega \in \Omega, \tilde{A}(\omega) = \tilde{B}(\omega)$
Fuzzy Subsethood: $\tilde{A} \tilde{\subseteq} \tilde{B} := \forall \omega \in \Omega, \tilde{A}(\omega) \leq \tilde{B}(\omega)$

Typically, we use $\sqcup = \vee, \sqcap = \wedge$, and $c(\mu) = 1 - \mu$, although it must always be kept in mind that there are many other possibilities. The extension principle can be observed again, in that for crisp sets, the classical set operations are recovered.

Proposition 7: Consider two crisp subsets $A, B \subseteq \Omega$, and let $\mu_{\tilde{C}} = \chi_A \sqcup \chi_B$, and $\mu_{\tilde{D}} = \chi_A \sqcap \chi_B$ for some general norm $\sqcap$ and conorm $\sqcup$. Then $\mu_{\tilde{C}} = \chi_{A \cup B}$ and $\mu_{\tilde{D}} = \chi_{A \cap B}$.

So again, we have the ideas necessary to cast fuzzy systems as a kind of GIT. As with classical sets, the basic objects are the points $\omega \in \Omega$, but now the compound objects are all the fuzzy subsets $\tilde{A} \tilde{\subseteq} \Omega$, and the valuation is into [0, 1] instead of $\{0, 1\}$. The operations on fuzzy sets are defined above.

So now we can introduce the measure of the information content of a fuzzy set. There are at least two important concepts here. First, we can consider the "size" of a fuzzy set much like that of a crisp set, in terms of its cardinality. In the fuzzy set case, this is simply

$$|\tilde{A}| := \sum_{\omega\in\Omega} \tilde{A}(\omega) ,$$

noting that in accordance with the extension principle, this fuzzy cardinality of a crisp set is thereby simply its cardinality.

We can also discuss the "fuzziness" of a fuzzy set, intuitively as how much a fuzzy set departs from being a crisp set, or in other words, some sense of "distance" between the fuzzy set and its complement [13]. The larger the distance, the "crisper" the fuzzy set. Using Z to denote this quantity, and recalling our fuzzy complement operator above, we have:

$$Z(\tilde{A}) := \sum_{\omega\in\Omega} |\tilde{A}(\omega) - c(\tilde{A}(\omega))|,$$

which, when $1 - \cdot$ is used for c, becomes:

$$Z(\tilde{A}) = \sum_{\omega\in\Omega} (1 - |2\tilde{A}(\omega) - 1|).$$

Finally, we note the presence of the extension principle everywhere. In particular, all of the classical set operations are recovered in the case of crisp sets, that is, where $\forall \omega \in \Omega, \tilde{A}(\omega) \in \{0,1\}$.

9.3.3.2 Fuzzy Logic

We saw in Section 9.2.1 that we can interpret the value of a characteristic function of a subset χ_A as the truth value of a proposition T_A, and in this way set theoretical operations are closely coupled to logical operations, to the extent of isomorphism. In classical predicate logic, a proposition A is a linguistic, or declarative, statement contained within the universe of discourse Ω, which can be identified as being a collection of elements in Ω which are strictly true or strictly false.

Thus, it is reasonable to take our concept of a fuzzy set's membership function $\mu_{\tilde{A}}$ and derive an isomorphic fuzzy logic, and indeed, this is what is available. In contrast to the classical case, a fuzzy logic proposition is a statement involving some concept without clearly defined boundaries. Linguistic statements that tend to express subjective ideas and that can be interpreted slightly differently by various individuals typically involve fuzzy propositions. Most natural language is fuzzy, in that it involves vague and imprecise terms. Assessments of people's preferences about colors, menus, or sizes, or expert opinions about the reliability of components, can be used as examples of fuzzy propositions.

So mathematically, we can regard a fuzzy subset $\tilde{A}$ as a fuzzy proposition, and denote $T_{\tilde{A}}(\omega) := \tilde{A}(\omega) \in [0,1]$ as the extent to which the statement "ω is $\tilde{A}$" is true. In turn, we can invoke the GIT operators analogously to fuzzy set theory to provide our fuzzy logic operators. In particular, for two fuzzy propositions $\tilde{A}$ and $\tilde{B}$ we have:

Negation: $T_{\neg\tilde{A}}(\omega) = c(\tilde{A}(\omega))$
Disjunction: $T_{\tilde{A}\text{ or }\tilde{B}}(\omega) = \tilde{A}(\omega) \sqcup \tilde{B}(\omega)$
Conjunction: $T_{\tilde{A}\text{ and }\tilde{B}}(\omega) = \tilde{A}(\omega) \sqcap \tilde{B}(\omega)$
Implication: There are actually a number of expressions for fuzzy implication available, but the "standard" one one might expect is valid: $T_{\tilde{A}\to\tilde{B}}(\omega) = T_{\neg\tilde{A}\text{ or }\tilde{B}}(\omega) = c(\tilde{A}(\omega)) \vee \tilde{B}(\omega)$

Table 9.4 shows the isomorphic relations among all the primary fuzzy operations. Again, the extension principle holds everywhere for crisp logic. Note, however, that, in keeping with the multi-valued nature of mathematical ideas in the more general theory, the implication operation is only roughly equivalent to the subset relation, and that there are other possibilities. Figure 9.3 shows our simple example again, along with the illustration of the generalization of classical sets and logic provided by fuzzy sets and logic.

TABLE 9.4 Isomorphisms between Fuzzy Logical and Fuzzy Set Theoretical Operations

Fuzzy Logic		Fuzzy Set Theory		GIT Operation
Negation	$\neg\tilde{A}$	Complement	$\tilde{A}^c$	$c(\mu_{\tilde{A}})$
Disjunction	$\tilde{A}$ or $\tilde{B}$	Union	$\tilde{A}\cup\tilde{B}$	$\mu_{\tilde{A}} \sqcup \mu_{\tilde{B}}$
Conjunction	$\tilde{A}$ and $\tilde{B}$	Intersection	$\tilde{A}\cap\tilde{B}$	$\mu_{\tilde{A}} \sqcap \mu_{\tilde{B}}$
Implication	$\tilde{A}\rightarrow\tilde{B}$	Subset	$\tilde{A}\subseteq\tilde{B}$	$c(\mu_{\tilde{A}}) \sqcup \mu_{\tilde{B}}$

9.3.3.3 Comparing Fuzzy Systems and Probability

The membership function $\tilde{A}(\omega)$ reflects an assessor's view of the extent to which $\omega \in \tilde{A}$, an epistemic uncertainty stemming from the lack of knowledge about how to classify ω. The subjective or personalistic interpretation of probability, $\Pr(A)$, can be interpreted as a two-sided bet, dealing with the uncertainty associated with the outcome of the experiment. While this type of uncertainty is usually labeled as random or aleatory, there is no restriction on applying subjective probability to characterize lack of knowledge or epistemic uncertainty. A common example would be eliciting probability estimates from experts for one-of-a kind or never observed events.

However, just because probabilities and fuzzy quantities can represent epistemic uncertainties does not guarantee interchangeability or even a connection between the two theories. As noted above, their axioms are quite different in how to combine uncertainties represented within each theory. Therefore, the linkage between the two theories is not possible by modifying one set of axioms to match the other. Other fundamental properties also differ. It is not a requirement that the sum over ω of all $\tilde{A}(\omega)$ equals one, as is required for summing over all probabilities. This precludes $\tilde{A}(\omega)$ from being interpreted as a probability in general. Similarly, pdfs are required to sum or integrate to one, but membership functions are not. Therefore, membership functions cannot be equated with pdfs either.

At least one similarity of probability and membership functions is evident. Just as probability theory does not tell how to specify $\Pr(A)$, fuzzy set theory does not tell how to specify $\tilde{A}(\omega)$. In addition, specifying membership is a subjective process. Therefore, subjective interpretation is an important common link to both theories.

Noting that $\tilde{A}(\omega)$, as a function of ω, reflects the extent to which $\omega \in \tilde{A}$, it is an indicator of *how likely* it is that $\omega \in \tilde{A}$. One interpretation of $\tilde{A}(\omega)$ is as the likelihood of ω for a fixed (specified) $\tilde{A}$. A likelihood function is not a pdf. In statistical inference, it is the relative degree of support that an observation provides to several hypotheses. Specifying the likelihood is also a subjective process, consistent with membership function definition and the subjective interpretation of probability.

As noted above, likelihoods are mostly commonly found in Bayes theorem. So, Bayes theorem links subjective probability with subjective likelihood. If membership functions can be interpreted as likelihoods, then Bayes theorem provides a valuable link from fuzzy sets back into probability theory. A case is made for this argument [30], providing an important mathematical linkage between probability and fuzzy theories. With two theories linked, it is possible to analyze two different kinds of uncertainties present in the same complex problem. An example application in the use of expert knowledge illustrates how these two theories can work in concert as envisioned by Zadeh [17] can be found in [31]. Additional research is needed to link other GITs so that different kinds of uncertainties can be accommodated within the same problem.

9.3.3.4 Fuzzy Arithmetic

Above we considered the restriction of sets from a general universe Ω to the line $\mathbb{R}$. Doing the same for fuzzy sets recovers some of the most important classes of structures.

In particular, we can define a **fuzzy quantity** as a fuzzy subset $\tilde{I} \tilde{\subseteq} \mathbb{R}$, such that $\mu_{\tilde{I}} : \mathbb{R} \mapsto [0,1]$. Note that a fuzzy quantity is any arbitrary fuzzy subset of $\mathbb{R}$, and as such may not have any particular useful properties. Also note that every pdf is a special kind of fuzzy quantity. In particular, if $\int_{\mathbb{R}} \mu_{\tilde{I}}(x)dx = 1$, then $\mu_{\tilde{I}}$ is a pdf.

We can discuss another special kind of fuzzy quantity, namely a **possibilistic density** or **distribution function** (again, depending on context), which we will abbreviate as π-df, pronounced "pie-dee-eff." In contrast with a pdf, if $\tilde{I}$ is fuzzy normal, so that $\sup_{x\in\mathbb{R}} \tilde{I}(x) = 1$, then $\tilde{I}$ is a π-df. We note here in passing that where a pdf is a special kind of probability distribution on $\mathbb{R}$, so a π-df is a special kind of possibility distribution on $\mathbb{R}$. This will be discussed in more detail below in Section 9.3.6.

We can also discuss special kinds of π-dfs. When a π-df is convex, so that

$$\forall_{x,y\in\mathbb{R}} \ \forall_{z\in[x,y]} \ \tilde{I}(z) \geq \tilde{I}(x) \wedge \tilde{I}(y), \tag{9.8}$$

then $\tilde{I}$ is a **fuzzy interval**. We can also define the **support** of a fuzzy interval as $\mathrm{U}(\tilde{I}) := \{x : \tilde{I}(x) > 0\}$, and note that $\mathbf{U}(\tilde{I})$ is itself a (possibly open) interval. When a fuzzy interval $\tilde{I}$ is unimodal, so that $\exists! x \in \mathbb{R}, \tilde{I}(x) = 1$, then $\tilde{I}$ is a **fuzzy number**, where $\exists!$ means "exists uniquely."

These classes of fuzzy quantities are illustrated in Figure 9.4. Note in particular that the fuzzy quantity and π-df illustrations are cartoons: in general, these need not be continuous, connected, or unimodal. Some of the cases shown here will be discussed further in Section 9.3.6.4.

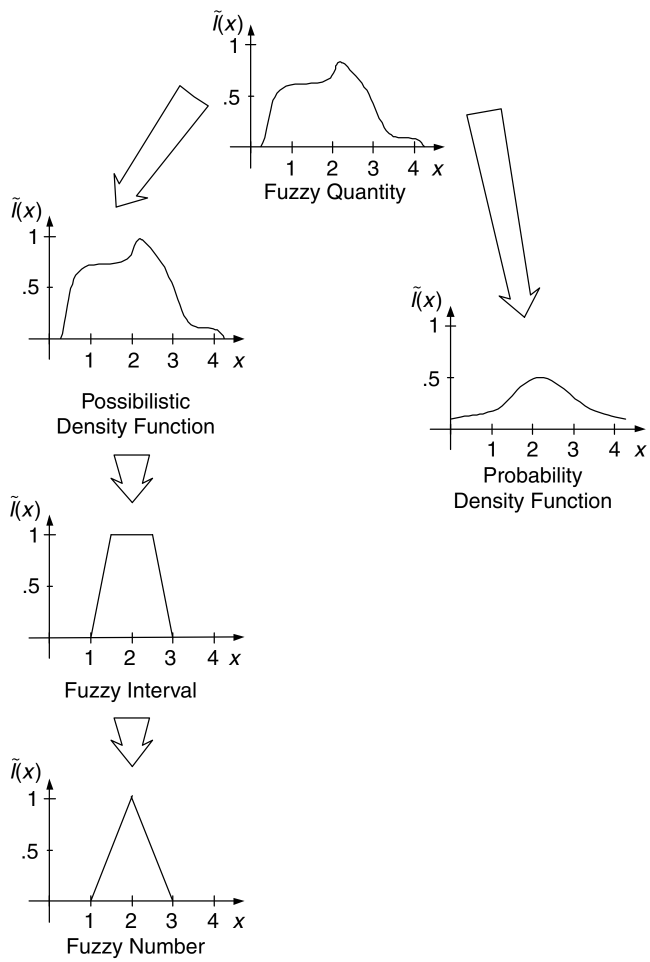

FIGURE 9.4 Kinds of fuzzy quantities.

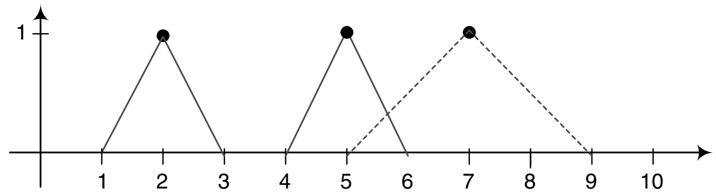

FIGURE 9.5 The fuzzy arithmetic operation [1, 2, 3] + [4, 5, 6] = [5, 7, 9].

Fuzzy intervals and numbers are named deliberately to invoke their extension from intervals and numbers. In particular, if a fuzzy interval $\tilde{I}$ is crisp, so that $\mu_{\tilde{I}} \mathbb{R} \mapsto \{0,1\}$, then $\tilde{I}$ is a crisp interval I with characteristic function $\chi_I = \mu_{\tilde{I}}$. Similarly, if a fuzzy number $\tilde{I}$ is crisp with mode x_0, so that $\tilde{I}(x_0) = 1$ and $\forall x \neq x_0, \tilde{I}(x) = 0$, then $\tilde{I}$ is just the number x_0, also characterized as the crisp interval $[x_0, x_0]$.

So this clears the way for us to define operations on fuzzy intervals, necessary to include it as a branch of GIT. As with crisp intervals, we are concerned with two fuzzy intervals $\tilde{I}, \tilde{J}$, and operations $* \in \{+, -, \times, \div\}$, etc. Then we have $\forall x \in \mathbb{R}$,

$$\mu_{\tilde{I} * \tilde{J}}(x) := \bigsqcup_{x = y * z} \left[\tilde{I}(y) \sqcap \tilde{J}(z) \right], \tag{9.9}$$

for some conorm $\sqcup$ and norm $\sqcap$. Again, the extension principle is adhered to, in that when $\tilde{I}$ and $\tilde{J}$ is crisp, Equation 9.3 is recovered from Equation 9.9.

An example of a fuzzy arithmetic operation is shown in Figure 9.5. We have two fuzzy numbers, each indicated by the triangles on the left. The leftmost, $\tilde{I}$, is unimodal around 2, and the rightmost, $\tilde{J}$, is unimodal around 6. Each is convex and normal, dropping to the *x*-axis as shown. Thus, $\tilde{I}$ expresses "about 2," and $\tilde{J}$ "about 5," and because they can be characterized by the three quantities of the mode and the *x*-intercepts, we denote them as

$$\tilde{I} = [1, 2, 3], \quad \tilde{J} = [4, 5, 6].$$

Applying Equation 9.9 for $* = +$ reveals $\tilde{I} + \tilde{J} = [5, 7, 9]$, which is "about 7."

Note how the extension principle is observed for fuzzy arithmetic as a generalization of interval arithmetic, in particular, we have

$$\mathbf{U}(\tilde{I} + \tilde{J}) = [5, 9] = [1, 3] + [4, 6] = \mathbf{U}(\tilde{I}) + \mathbf{U}(\tilde{J}),$$

where the final operation indicates interval arithmetic as in Equation 9.3.

The relations among these classes of fuzzy quantities, along with representative examples of each, is shown in Figure 9.6

9.3.3.5 Interpretations and Applications

Some simple examples can illustrate the uncertainty concept and construction of a fuzzy set, and the corresponding membership function.

First, let Ω be the set of integers between zero and ten, inclusive: $\Omega = \{0, 1, 2, \ldots, 10\}$. Suppose we are interested in a subset of Ω, $\tilde{A}$, where $\tilde{A}$ contains all the *medium* integers of Ω: $\tilde{A} = \{\omega : \omega \in \Omega$ and ω is medium$\}$. To specify $\tilde{A}$, the term "medium integer" must be defined. Most would consider 5 as medium, but what about 7? The uncertainty (or vagueness) about what constitutes a medium integer is what makes $\tilde{A}$ a fuzzy set, and such sets occur in our everyday use (or natural language). The uncertainty of

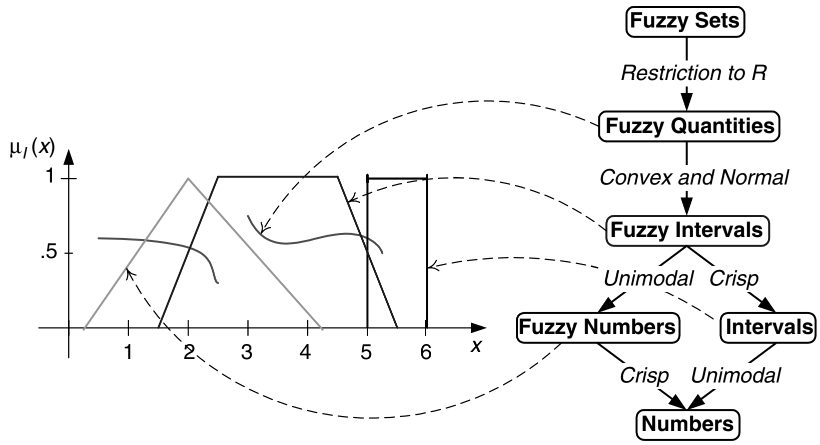

FIGURE 9.6 Fuzzy quantities.

classification arises because the boundaries of $\tilde{A}$ are not crisp. The integer 7 might have some membership (belonging) in $\tilde{A}$ and yet also have some degree of membership in $\tilde{A}^c$. Said another way, the integer 7 might have some membership in $\tilde{A}$ and yet also have some membership in another fuzzy set, $\tilde{B}$, where $\tilde{B}$ is the fuzzy set of large integers in Ω.

For a more meaningful example, assume we have a concept design for a new automotive system, like a fuel injector. Many of its components are also new designs, but may be similar to ones used in the past, implying that partial knowledge exists that is relevant to the new parts, but also implying large uncertainties exist about the performance of these parts and the system. The designer of this system wants to assess its performance based upon whatever information is currently available before building prototypes or implementing expensive test programs. The designer also wants to be assured that the performance is "excellent" with a high confidence. This desire defines a reliability **linguistic variable**. While reliability is traditionally defined as the probability that a system performs its functions for a given period of time and for given specifications, the knowledge about performance (especially new concepts) may only be in the form of linguistic and fuzzy terms.

For example, a component designer may only have access to the information that "if the temperature is too hot, this component won't work very well." The conditions (e.g., "too hot") can be characterized by a fuzzy set, and the performance (e.g., "won't work very well") can also be represented by a fuzzy set. Chapter 11 of Ross, Booker, and Parkinson [15] illustrates how fuzzy sets can be used for linguistic information and then combined with test data, whose uncertainty is probabilistic, to form a traditionally defined reliability.

Combining the probabilistic uncertainty of outcomes of tests and uncertainties of fuzzy classification from linguistic knowledge about performance requires a theoretical development for linking the two theories. Linkage between the probability and fuzzy set theories can be accomplished through the use of Bayes theorem, whose two ingredients are a prior probability distribution function and a likelihood function. As discussed in Section 9.3.3.3, Singpurwalla and Booker [30] relax the convention that the maximum value of $\tilde{A}(\omega)$ is set to 1.0, because that better conforms to the definition of a likelihood. Their theoretical development demonstrates the equivalency of likelihood and membership.

In the example above, if test data exists on a component similar to a new concept design component then probability theory could be used to capture the uncertainties associated with that data set, forming the prior distribution in Bayes theorem. Expert knowledge about the new design in the form of linguistic information about performance could be quantified using fuzzy membership functions, forming the likelihood. The combination of these two through Bayes theorem produces a posterior distribution, providing a probability based interpretation of reliability for the component. See [15] for more details on this kind of approach.

9.3.4 Monotone and Fuzzy Measures

In discussing probability theory in Section 9.2.3, we distinguished the probability measure Pr valued on sets $A \subseteq \Omega$ from the probability distribution p valued on points $\omega \in \Omega$. Then in Section 9.3.3 we characterized the membership functions of fuzzy sets $\tilde{A}$ also as being valued on points $\omega \in \Omega$, and, indeed, that probability distributions and pdfs are, in fact, kinds of fuzzy sets. It is natural to consider classes of measures other than Pr which are also valued on subsets $A \subseteq \Omega$, and perhaps related to other kinds of fuzzy sets.

This is the spirit that inspired Sugeno to define classes of functions he called **fuzzy measures** [25, 32]. Since then, terminological clarity has led us to call these **monotone measures** [33].

Assume for the moment a finite universe of discourse Ω, and then define a monotone measure as a function $\nu : 2^{\Omega} \mapsto [0,1]$, where $\nu(\varnothing) = 0, \nu(\Omega) = 1$, and

$$A \subseteq B \rightarrow \nu(A) \leq \nu(B). \tag{9.10}$$

When Ω is uncountably infinite, continuity requirements on ν come into play, but this will suffice for us for now.

We can also define the **trace** of a monotone measure as the generalization of the concept of a density or distribution. For any monotone measure ν, define its trace as a function $\rho_\nu : \Omega \mapsto [0,1]$, where $\rho_\nu(\omega) := \nu(\{\omega\})$.

In general, measures are much "larger" than traces, in that they are valued on the space of subsets $A \subseteq \Omega$, rather than the space of the points of $\omega \in \Omega$. So, for finite Ω with $n = |\Omega|$, a trace needs to be valued n times, one for each point $\omega \in \Omega$, while a measure needs to be valued 2^n times, one for each subset $A \subseteq \Omega$. Therefore, it is very valuable to know if, for a particular measure, it might be possible not to know all 2^n values of the measure independently, but rather to be able to calculate some of these based on knowledge of the others; in other words, to be able to break the measure into a small number of pieces and then put those pieces back together again. This greatly simplifies calculations, visualization, and elicitation.

When this is the case, we call such a monotone measure **distributional** or **decomposable**. Mathematically, this is the case when there exists a conorm $\sqcup$ such that

$$\forall A, B \subseteq \Omega, \quad \nu(A \cup B) + \nu(A \cap B) = \nu(A) \sqcup \nu(B). \tag{9.11}$$

It follows that

$$\nu(A \cup B) = \nu(A) \sqcup \nu(B) - \nu(A \cap B).$$

It also follows that when $A \cap B = \varnothing$, then $\nu(A \cup B) = \nu(A) \sqcup \nu(B)$.

Decomposability expresses the idea that the measure can be broken into pieces. The smallest such pieces are just the values of the trace, and thus decomposability is also called "distributionality," and can be expressed as

$$\nu(A) = \bigsqcup_{\omega \in A} \rho_\nu(\omega).$$

Finally, we call a monotone measure **normal** when $\nu(\Omega) = 1$. When ν is both normal and decomposable, it follows that

$$\bigsqcup_{\omega \in \Omega} \rho_\nu(\omega) = 1.$$

So it is clear that every probability measure Pr is a monotone measure, but not vice versa, and the distribution p of a finite probability measure and the pdf f of a probability measure on $\mathbb{R}$ are both traces of the corresponding measure Pr. Indeed, all of these concepts are familiar to us from probability theory, and are, in fact, direct generalizations of it.

In particular, a probability measure Pr is a normal, monotone measure that is decomposable for the bounded sum conorm $+_b$ (see Section 9.3.2), and whose trace is just the density. Note, however, that because we presume that a probability measure Pr is always normalized, when operating on probability values, the bounded sum conorm $+_b$ becomes equivalent to addition +. For example, when $\sum_{\omega\in\Omega} p(\omega)=1$, then $\forall \omega_1,\omega_2 \in \Omega, p(\omega_1)+_b p(\omega_2)=p(\omega_1)+p(\omega_2)$. In this way we recover the familiar results for probability theory:

$$\begin{gathered}
\Pr(A\cup B)=\Pr(A)+\Pr(B)-\Pr(A\cap B) \\
A\cap B=\varnothing \ \rightarrow \Pr(A\cup B)=\Pr(A)+\Pr(B) \\
\Pr(A)+\Pr(A^c)=1 \\
\Pr(A)=\sum_{\omega\in A} p(\omega).
\end{gathered} \tag{9.12}$$

Note that a trace $\rho_v:\Omega \mapsto [0,1]$ is a function to the unit interval, and is thus a fuzzy set. This will be important below, as in many instances it is desirable to interpret the traces of fuzzy measures such as probability or possibility distributions as special kinds of fuzzy sets.

A measure of information content in the context of general fuzzy or monotone measures is an area of active research, and beyond the scope of this chapter (see elsewhere for details [33]). Below we consider some particular cases in the context of random sets and possibility theory.

9.3.5 Random Sets and Dempster-Shafer Evidence Theory

In our historical discussion in Section 9.3.1, we noted that one of the strongest threads in GIT dates back to Dempster's work in probability measures propagated through multi-valued maps, and the subsequent connection to Shafer's theory of evidence and random sets. We detail this in this section, building on the ideas of monotone measures.

9.3.5.1 Dempster-Shafer Evidence Theory

In particular, we can identify **belief** Bel and **plausibility** Pl as dual fuzzy measures with the properties of super- and sub-additivity, respectively:

$$\begin{gathered}
\mathrm{Bel}(A\cup B)\geq \mathrm{Bel}(A)+\mathrm{Bel}(B)-\mathrm{Bel}(A\cap B) \\
\mathrm{Pl}(A\cap B)\leq \mathrm{Pl}(A)+\mathrm{Pl}(B)-\mathrm{Pl}(A\cup B).
\end{gathered}$$

Note the contrast with the additivity of a probability measure shown in Equation 9.12. In particular, it follows that each probability measure Pr is both a belief and a plausibility measure simultaneously. Also, while Pr is always decomposable in $+_b$, Bel and Pl are only decomposable under some special circumstances.

Again, in contrast with probability, we have the following sub- and super-additive properties for Bel and Pl:

$$\mathrm{Bel}(A)+\mathrm{Bel}(A^c)\leq 1, \qquad \mathrm{Pl}(A)+\mathrm{Pl}(A^c)\geq 1$$

Also, Bel and Pl are dually related, with:

$$\mathrm{Bel}(A)\leq \mathrm{Pl}(A), \tag{9.13}$$

$$\mathrm{Bel}(A)=1-\mathrm{Pl}(A^c), \quad \mathrm{Pl}(A)=1-\mathrm{Bel}(A^c). \tag{9.14}$$

So, not only are Bel and Pl co-determining, but each also determines and is determined by another function called a **basic probability assignment** $m: 2^{\Omega} \mapsto [0,1]$ where $m(\emptyset) = 0$, and

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{9.15}$$

$m(A)$ is also sometimes called the "mass" of A.

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B), \qquad \mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

We then have the following relations:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathrm{Bel}(B) = \sum_{B \subseteq A} (-1)^{|A-B|} (1 - \mathrm{Pl}(B^c)), \tag{9.16}$$

where Equation 9.16 expresses what is called a **Möbius inversion**. Thus, given any one of m, Bel, or Pl, the other two are determined accordingly.

Some other important concepts are:

- A **focal element** is a subset $A \subseteq \Omega$ such that $m(A) > 0$. In this chapter, we always presume that there are only a finite number N of such focal elements, and so we use the notation $A_j, 1 \leq j \leq N$ for all such focal elements.
- The **focal set** $\mathcal{F}$ is the collection of all focal elements:

$$\mathcal{F} = \{A_j \subseteq \Omega : m(A_j) > 0\}.$$

- The **support** of the focal set is the global union:

$$\mathbf{U} = \bigcup_{A_j \in \mathcal{F}} A_j.$$

- The **core** of the focal set is the global intersection:

$$\mathbf{C} = \bigcap_{A_j \in \mathcal{F}} A_j.$$

- A **body of evidence** is the combination of the focal set with their masses:

$$\mathcal{S} = \langle \mathcal{F}, m \rangle = \langle \{A_j\}, \{m(A_j)\} \rangle, \qquad 1 \leq j \leq N.$$

- Given two independent bodies of evidence $\mathcal{S}_1 = \langle \mathcal{F}_1, m_1 \rangle, \mathcal{S}_2 = \langle \mathcal{F}_2, m_2 \rangle$, then we can use **Dempster combination** to produce a combined body of evidence $\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_2 = \langle \mathcal{F}, m \rangle$, where $\forall A \subseteq \Omega$,

$$m(A) = \frac{\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)}{\sum_{A_1 \cap A_2 \neq \emptyset} m_1(A_1) m_2(A_2)}.$$

While Dempster's rule is the most prominent combination rule, there are a number of others available [34].

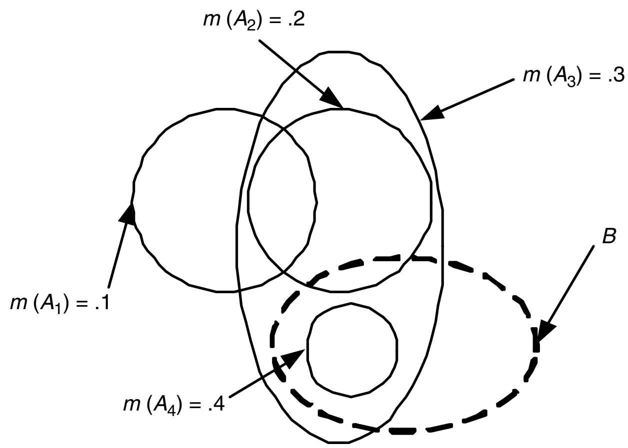

FIGURE 9.7 A Dempster-Shafer body of evidence.

- Assume a body of evidence $\mathcal{S}$ drawn from a finite universe of discourse with $\Omega = \{\omega_i\}$ with $1 \le i \le n$. Then if $N = n$, so that the number of focal elements is equal to the number of elements of the universe of discourse, then we call $\mathcal{S}$ **complete** [35].

An example is shown in Figure 9.7. We have:

$$\mathcal{F} = \{A_1, A_2, A_3, A_4\}$$

$$m(A_1) = .1, \quad m(A_2) = .2, \quad m(A_3) = .3, \quad m(A_4) = .4$$

$$\mathrm{Bel}(B) = \sum_{A_j \subseteq B} m(A_j) = m(A_4) = .4$$

$$\mathrm{Pl}(B) = \sum_{A_j \cap B \neq \emptyset} m(A_j) = m(A_2) + m(A_3) + m(A_4) = .2 + .3 + .4 = .9.$$

9.3.5.2 Random Sets

Above we identified the structure $\mathcal{S} = \langle \mathcal{F}, m \rangle = \langle \{A_j\}, \{m(A_j)\} \rangle$ as a body of evidence. When Ω is finite, we can express this body of evidence in an alternative form: instead of a pair of sets ($\mathcal{F}$ and m), now we have a set of pairs, in particular, pairings of focal elements $A_j \in \mathcal{F}$ with their basic probability value $m(A_j)$:

$$\langle \{A_j\}, \{m(A_j)\} \rangle \mapsto \{\langle A_j, m(A_j) \rangle\}.$$

We call this form the **random set** representation of the DS body of evidence.

This alternative formulation is triggered by recalling that $\sum_{A_j \in \mathcal{F}} m(A_j) = 1$, so that m can be taken as a discrete probability distribution or density on the various sets A_j. In other words, we can interpret $m(A_j)$ as the probability that A_j occurs compared to all the other $A \subseteq \Omega$.

Note that despite a superficial similarity, there is a profound difference between a probability measure $\Pr(A)$ and a basic probability assignment $m(A)$. Where it must *always* be the case that for two sets $A, B \subseteq \Omega$, Equation 9.12 must hold, in general there need be *no* relation between $m(A)$ and $m(B)$, other than Equation 9.15, that $\sum_{A \subseteq \Omega} m(A) = 1$.

To explicate this difference, we recall the definition of a **random variable**. Given a probability space $\langle \Omega, \mathcal{E}, \Pr \rangle$, then a function $S : X \mapsto \Omega$ is a random variable if S is "Pr-measurable," so that $\forall \omega \in \Omega$,

$S^{-1}(\omega) \in \mathcal{E}$. S then assigns probabilities to the items $\omega \in \Omega$. Similarly, we can think of a random set simply as a random variable that takes values on *collections* or *sets* of items, rather than points.

General Random Set: $\mathcal{S}: X \mapsto 2^{\Omega} - \{\emptyset\}$ is a random subset of Ω if $\mathcal{S}$ is Pr-measurable: $\forall \emptyset \neq A \subseteq \Omega$ $\mathcal{S}^{-1}(A) \in \Sigma$. m acts as a density of $\mathcal{S}$.

Given this, we can then interpret the DS measures in a very natural way in terms of a random set $\mathcal{S}$:

$$m(A) = \Pr(\mathcal{S} = A),$$

$$\mathrm{Bel}(A) = \Pr(\mathcal{S} \subseteq A),$$

$$\mathrm{Pl}(A) = \Pr(\mathcal{S} \cap A \neq \emptyset).$$

In the remainder of the chapter we will generally refer to random sets, by which we will mean finite random sets, which are isomorphic to finite DS bodies of evidence. Also, for technical reasons (some noted below in Section 9.3.6.2), there is a tendency to work only with the plausibility measure Pl, and to recover the belief Bel simply by the duality relation (Equation 9.14). In particular, for a general random set, we will generally consider its trace specifically as the plausibilistic trace ρ_{Pl}, and thereby have

$$\rho_{\mathrm{Pl}}(\omega_i) = \mathrm{Pl}(\{\omega_i\}) = \sum_{A_j \ni \omega_i} m_j.$$

The components of random sets as a GIT are now apparent. The basic components now are not points $\omega \in \Omega$, but rather subsets $A \subseteq \Omega$, and the compound objects are the random sets $\mathcal{S}$. The valuation set is again [0, 1], and the valuation is in terms of the evidence function m. Finally, operations are in terms of the kinds of combination rules discussed above [34], and operations defined elsewhere, such as inclusion of random sets [36].

9.3.5.3 The Information Content of a Random Set

The final component of random sets as a GIT, namely the measure of the information content of a random set, has been the subject of considerable research. Development of such a measure is complicated by the fact that random sets by their nature incorporate two distinct kinds of uncertainty. First, because they are random variables, they have a probabilistic component best measured by entropies, as in Equation 9.6. But, unlike pure random variables, their fundamental "atomic units" of variation are the focal elements A_j, which differ from each other in size $|A_j|$ and structure, in that some of them might overlap with each other to one extent or another. These aspects are more related to simple sets or intervals, and thus are best measured by measures of "nonspecificity" such as the Hartley measure of Equations 9.1 and 9.4.

Space precludes a discussion of the details of developing information measure for random sets (see [33]). The mathematical development has been long and difficult, but we can describe some of the highlights here.

The first good candidate for a measure of uncertainty in random sets is given by generalizing the nonspecificity of Equations 9.1 and 9.4 to be:

$$U_{\mathbf{N}}(\mathcal{S}) := -\sum_{A_j \in \mathcal{F}} m_j \log_2(|A_j|). \tag{9.17}$$

This has a number of interpretations, the simplest being the expectation of the size of a focal element. Thus, both components of uncertainty are captured: the randomness of the probabilistic variable coupled to the variable size of the focal element.

While this nonspecificity measure $U_{\mathbf{N}}$ captures many aspects of uncertainty in random sets, it does not as well capture all the attributes related to conflicting information in the probabilistic component m, which is reflected in probability theory by the entropy of Equation 9.6. A number of measures have been suggested, including **conflict** as the entropy of the singletons:

$$U_S(\mathcal{S}) := -\sum_{\omega_i \in \Omega} m(\{\omega_i\}) \log_2(m(\{\omega_i\})),$$

and **strife** as a measure of entropy focused on individual focal elements:

$$U_S(\mathcal{S}) := -\sum_{A_j \in \mathcal{F}} m_j \log_2\left[\sum_{k=1}^{N} m_k \frac{|A_j \cap A_k|}{|A_j|}\right].$$

While each of these measures can have significant utility in their own right, and arise as components of a more detailed mathematical theory, in the end, none of the them alone proved completely successful in the context of a rigorous mathematical development. Instead, attention has turned to single measures that attempt to directly integrate both nonspecificity and conflict information. These measures are not characterized by closed algebraic forms, but rather as optimization problems over sets of probability distributions. The simplest expression of these is given as an **aggregate uncertainty**:

$$U_{AU}(\mathcal{S}) := \max_{p: \forall A \subseteq \Omega, \Pr(A) \le \mathrm{Pl}(A)} U_{\text{prob}}(p), \tag{9.18}$$

recalling that $\Pr(A) = \sum_{\omega_i \in A} p(\omega_i)$ and $U_{\text{prob}}(p)$ is the statistical entropy (Equation 9.6). In English, U_{AU} is the largest entropy of all probability distributions consistent with the random set $\mathcal{S}$.

9.3.5.4 Specific Random Sets and the Extension Principle

The extension principle also holds for random sets, recovering ordinary random variables in a special case. So, given that random sets are set-valued random variables, then we can consider the special case where each focal element is not, in fact, a set at all, but really just a point, in particular a singleton set. We call such a focal set **specific**:

$$\forall A_j \in \mathcal{F}, \quad |A_j| = 1, \quad \exists! \omega_i \in \Omega, A_j = \{\omega_i\}.$$

When a specific random set is also complete, then conversely we have that $\forall \omega_i \in \Omega, \exists! A_j \in \mathcal{F}, A_j = \{\omega_i\}$.

Under these conditions, the gap between $\mathrm{Pl}(A)$ and $\mathrm{Bel}(A)$ noted in Equation 9.13 closes, and this common DS measure is just a probability measure again:

$$\forall A \subseteq \Omega, \quad \mathrm{Pl}(A) = \mathrm{Bel}(A) = \Pr(A). \tag{9.19}$$

And when $\mathcal{S}$ is complete, the plausibilistic trace reverts to a probability distribution, with $p(\omega_i) = m_j$ for that A_j which equals $\{\omega_i\}$.

As an example, consider Figure 9.8, where $\Omega = \{x, y, z, w\}$, and $\mathcal{F} = \{\{x\}, \{y\}, \{z\}, \{w\}\}$ with m as shown. Then, for $B = \{z, w\}, C = \{y, w\}$, we have

$$\mathrm{Bel}(B) = \mathrm{Pl}(B) = \Pr(B) = .3 + .4 = .7, \quad \Pr(B \cup C) = \Pr(B) + \Pr(C) - \Pr(B \cap C) = .9,$$

and, using vector notation, the probability distribution is $p = \langle .1, .2, .3, .4 \rangle$.

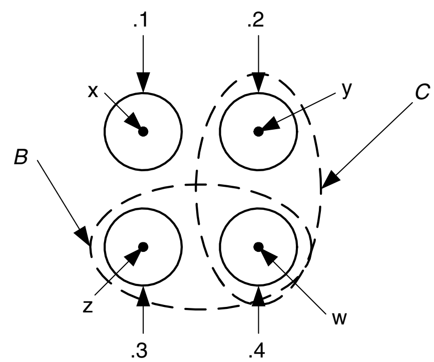

FIGURE 9.8 A specific random set, which induces a probability distribution.

9.3.5.5 Random Intervals and P-Boxes

Above we moved from sets to intervals, and then fuzzy sets to fuzzy intervals. Now we want to similarly move from random sets to **random intervals**, or DS structures on the Borel field $\mathcal{D}$ defined in Equation 9.2. Define a random interval $\mathcal{A}$ as a random set on $\Omega = \mathbb{R}$ for which $\mathcal{F}(\mathcal{A}) \subseteq \mathcal{D}$. Thus, a random interval is a random left-closed interval subset of $\mathbb{R}$. For a random interval, we denote the focal elements as intervals $I_j, 1 \leq j \leq N$, so that $\mathcal{F}(\mathcal{A}) = \{I_j\}$.

An example is shown in Figure 9.9, with $N = 4$,

$$\mathcal{F} = \{[2.5, 4), [1, 2), [3, 4), [2, 3.5)\},$$

support $\mathbf{U}(\mathcal{F}(\mathcal{A})) = [1, 4)$, and m is as shown.

Random interval approaches are an emerging technology for engineering reliability analysis [37–41]. Their great advantage is their ability to represent not only randomness via probability theory, but also imprecision and nonspecificity via intervals, in an overall mathematical structure that is close to optimally simple. As such, they are superb ways for engineering modelers to approach the world of GIT.

So, random intervals are examples of DS structures restricted to intervals. This restriction is very important because it cuts down substantially on both the quantity of information and computational and human complexity necessary to use such structures for modeling. Even so, they remain relatively complex structures, and can present challenges to modelers and investigators in their elicitation and interpretation. In particular, interpreting the fundamental structures such as possibly overlapping focal elements and basic probability weights can be a daunting task for the content expert, and it can be desirable to interact with investigators over more familiar mathematical objects. For these reasons, we commonly introduce simpler mathematical structures that approximate the complete random interval by representing a portion of their information. We introduce these now.

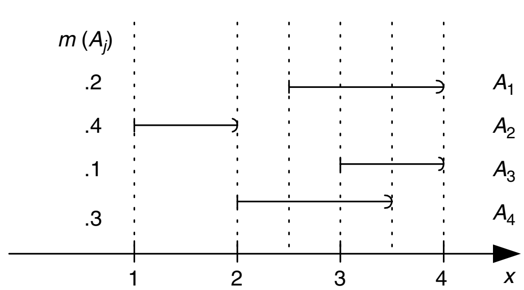

FIGURE 9.9 A random interval.

A probability box, or just a **p-box** [42], is a structure $\mathcal{B} := \langle \underline{B}, \overline{B} \rangle$, where $\underline{B}, \overline{B} : \mathbb{R} \mapsto [0,1]$,

$$\lim_{x \mapsto -\infty} B(x) \to 0, \quad \lim_{x \mapsto \infty} B(x) \to 1, \quad B \in \mathcal{B},$$

and $\underline{B}, \overline{B}$ are monotonic with $\underline{B} \leq \overline{B}$. $\underline{B}$ and $\overline{B}$ are interpreted as bounds on cumulative distribution functions (CDFs). In other words, given $\mathcal{B} = \langle \underline{B}, \overline{B} \rangle$, we can identify the set of all functions $\{F : \underline{B} \leq F \leq \overline{B}\}$ such that F is the CDF of some probability measures Pr on $\mathbb{R}$. Thus each p-box defines such a class of probability measures.

Given a random interval $\mathcal{A}$, then

$$\mathcal{B}(\mathcal{A}) := \langle \text{BEL, PL} \rangle \tag{9.20}$$

is a p-box, where BEL and PL are the "cumulative belief and plausibility distributions" $\text{PL}, \text{BEL} : \mathbb{R} \mapsto [0, 1]$ originally defined by Yager [43]

$$\text{BEL}(x) := \text{Bel}((-\infty, x)), \quad \text{PL}(x) := \text{Pl}((-\infty, x)).$$

Given a random interval $\mathcal{A}$, then it is also valuable to work with its plausibilistic trace (usually just identified as its trace), where $r_{\mathcal{A}}(x) := \text{Pl}(\{x\})$. Given a random interval $\mathcal{A}$, then we also have that $r_{\mathcal{A}} = \text{PL} - \text{BEL}$, so that for a p-box derived from Equation 9.20, we have

$$r_{\text{A}} = \overline{B} - \underline{B}. \tag{9.21}$$

See details elsewhere [44, 45].

The p-box generated from the example random interval is shown in the top of Figure 9.10. Because $\overline{B}$ and $\underline{B}$ partially overlap, the diagram is somewhat ambiguous on its far left and right portions, but note that

$$\overline{B}((-\infty, 1)) = 0, \quad \underline{B}((-\infty, 2,)) = 0,$$
$$\overline{B}([3, \infty)) = 1, \quad \underline{B}([3.5, \infty)) = 1.$$

The trace $r_{\mathcal{A}} = \overline{B} - \underline{B}$ is also shown.

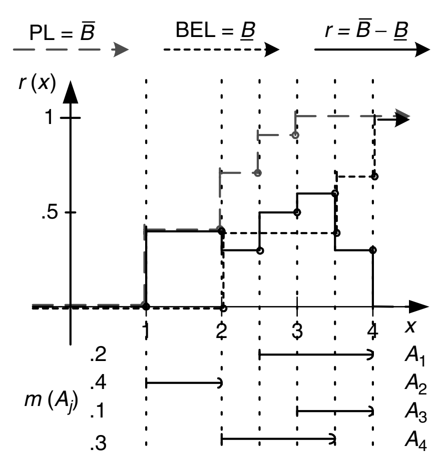

FIGURE 9.10 The probability box derived from a random interval.

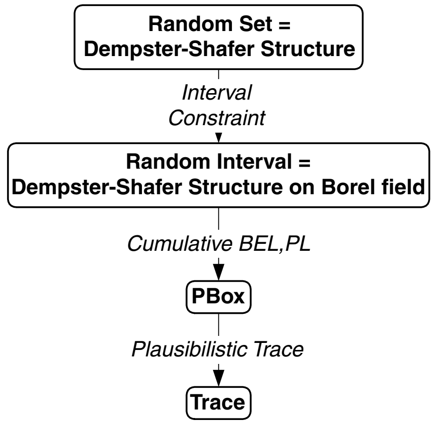

FIGURE 9.11 Relations among random sets, random intervals, p-boxes, and their traces.

So each random interval determines a p-box by Equation 9.20, which in turn determines a trace by Equation 9.21. But conversely, each trace determines an equivalence class of p-boxes, and each p-box an equivalence class of random intervals. In turn, each such equivalence class has a canonical member constructed by a standard mechanism. These relations are diagrammed in Figure 9.11, and see details elsewhere [45].

9.3.6 Possibility Theory

So far, we have discussed classical uncertainty theories in the form of intervals and probability distributions; their generalization to fuzzy sets and intervals; and the corresponding generalization to random sets and intervals. In this section we introduce **possibility theory** as a form of information theory, which in many ways exists as an alternative to and in parallel with probability theory, and which arises in the context of both fuzzy systems and DS theory.

9.3.6.1 Possibility Measures and Distributions

In Section 9.3.4 we identified a probability measure Pr as a normal monotone measure that is decomposable for the bounded sum conorm $+_b$. Similarly, a **possibility measure** Π is a normal, monotone measure that is decomposable for the *maximum* conorm $\vee$. In this way, the familiar results for probability theory shown in Equation 9.12 are replaced by their maximal counterparts for possibility theory. In particular, Equation 9.11 yields

$$\forall A, B \subseteq \Omega, \quad \Pi(A \cup B) \vee \Pi(A \cap B) = \Pi(A) \vee \Pi(B), \tag{9.22}$$

so that from Equation 9.10 it follows that

$$\Pi(A \cup B) = \Pi(A) \vee \Pi(B), \tag{9.23}$$

whether A and B are disjoint or not.

The trace of a possibility measure is called a **possibility distribution** $\pi : \Omega \mapsto [0,1], \pi(\omega) = \Pi(\{\omega\})$. Continuing our development, we have the parallel results from probability theory:

$$\Pi(A) = \bigvee_{\omega \in A} \pi(\omega), \quad \Pi(\Omega) = \bigvee_{\omega \in \Omega} \pi(\omega) = 1.$$

In Section 9.3.5.4 and Figure 9.8 we discussed how the extension principle recovers a "regular" probability measure from a random set when the focal elements are specific, so that the duality of belief and plausibility collapses and we have that $\mathrm{Bel} = \mathrm{Pl} = \mathrm{Pr}$. In a sense, possibility theory represents the opposite extreme case. In a specific random set, all the focal elements are singletons, and are thus maximally small and disjoint from each other. Possibility theory arises in the alternate case, when the focal elements are maximally large and intersecting.

In particular, we call a focal set **consonant** when $\mathcal{F}$ is a nested class, so that $\forall A_1, A_2 \in \mathcal{F}$, either $A_1 \subseteq A_2$ or $A_2 \subseteq A_1$. We can then arbitrarily order the A_j so that $A_i \subseteq A_{i+1}$, and we assign $A_0 := \emptyset$. If a consonant random set is also complete, then we can use the notation

$$\forall 1 \leq i \leq n, \quad A_i = \{\omega_1, \omega_2, \ldots, \omega_i\} \tag{9.24}$$

and we have that $A_i - A_{i-1} = \{\omega_i\}$.

Whenever a random set is consonant, then the plausibility measure Pl becomes a measure Π. But unlike in a specific random set, where the belief and plausibility become maximally close, in a consonant random set the belief and plausibility become maximally separated, to the point that we call the dual belief measure Bel a **necessity measure**:

$$\eta(A) = 1 - \Pi(A^c).$$

Where a possibility measure Π is characterized by the maximum property by Equation 9.23, a necessity measure η is characterized by

$$\eta(A \cap B) = \eta(A) \wedge \eta(B) \tag{9.25}$$

For the possibility distribution, when $\mathcal{S}$ is complete, and using the notation from Equation 9.24, we have:

$$\pi(\omega_i) = \sum_{j=i}^{n} m_j, \quad m(A_i) = \pi(\omega_i) - \pi(\omega_{i+1}),$$

where $\pi_{n+1} := 0$ by convention.

An example of a consonant random set is shown in Figure 9.12, where again $\Omega = \{x, y, z, w\}$, but now

$$\mathcal{F} = \{\{x\}, \{x, y\}, \{x, y, z\}, \{x, y, z, w\}\}$$

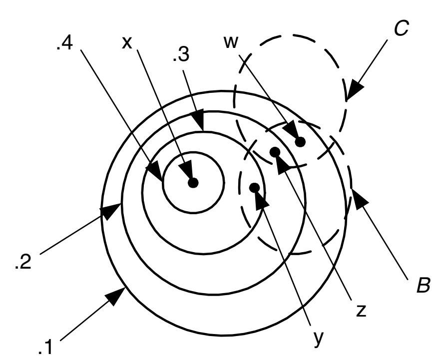

FIGURE 9.12 A consonant, possibilistic random set.

with m as shown. Then, for $B = \{y, z, w\}, C = \{z, w\}$, we have

$$\text{Pl}(B) = .1 + .2 + .3 = .6, \quad \text{Pl}(C) = .1 + .2 = .3,$$
$$\text{Pl}(B \cup C) = .1 + .2 + .3 = .6 = \text{Pl}(B) \vee \text{Pl}(C),$$

thus characterizing Pl as, in fact, a possibility measure Π. We also have, using vector notation, $\pi = \langle 1, .6, .3, .1 \rangle$.

In considering the information content of a possibility measure, our intuition tells us that it would best be thought of as a kind of nonspecificity such as in Equation 9.17. However, while the nonspecificity of a specific random set vanishes, the strife or conflict of a consonant random set does not. Indeed, it was exactly this observation that drove much of the mathematical development in this area, and thus, technically, the information content of a possibility measure is best captured by an aggregate uncertainty such as Equation 9.18.

However, for our purposes, it is useful to consider Equation 9.17 applied to consonant random sets. Under these conditions, we can express Equation 9.17 in terms of the possibility distribution as:

$$U_{\text{N}}(\pi) := \sum_{i=2}^{n} \pi_i \log_2\left(\frac{i}{i-1}\right) = \sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2(i).$$

9.3.6.2 Crispness, Consistency, and Possibilistic Histograms

In the case of the specific random set discussed in Section 9.3.5.4, not only do the belief measure Bel and plausibility measure Pl collapse together to the decomposable probability measure Pr, but also their traces ρ_{Bel} and ρ_{Pl} collapse to the trace of the probability measure ρ_{Pr}, which is just the probability density p.

But the possibilistic case, which is apparently so parallel, also has some definite differences. First, as we saw, the belief and plausibility measures are distinct as possibility Π and necessity η. Moreover, the necessity measure η is not decomposable, and indeed, the minimum operator $\wedge$ from Equation 9.25 is not a conorm. But moreover, the relation between the measure Pl and its trace ρ_{Pl} is also not so simple.

First consider the relation between intervals and fuzzy intervals discussed in Section 9.3.3.4. In particular, regular (crisp) intervals arise through the extension principle when the characteristic function takes values only in $\{0, 1\}$. Similarly, for general possibility distributions, it might be the case that $\pi(\omega_i) \in \{0, 1\}$. In this case, we call π a **crisp possibility distribution**, and otherwise identify a noncrisp possibility distribution as a **proper possibility distribution**. Thus, each fuzzy interval is a general (that is, potentially proper) possibility distribution, while each crisp interval is correspondingly a crisp possibility distribution.

Note that crispness can arise in probability theory only in a degenerate case, because if $\exists \omega_i, p(\omega_i) = 1$, then for all other values, we would have $p(\omega_i) = 0$. We call this case a **certain distribution**, which are the only cases that are both probability and possibility distributions simultaneously.

Now consider random sets where the core is non-empty, which we call **consistent**:

$$\text{C} = \bigcap_{A_j \in \mathcal{F}} A_j \neq \emptyset.$$

Note that all consistent random sets are consonant, but not *vice versa*. But in these cases, the trace ρ_{Pl} is a maximal possibility distribution satisfying

$$\bigvee_{\omega \in \Omega} \rho_{\text{Pl}}(\omega) = 1,$$

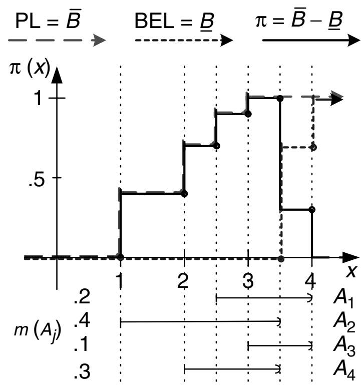

FIGURE 9.13 A consistent random interval with its possibilistic histogram π.

but Pl is not a possibility measure (that is, Equation 9.22 is *not* satisfied). However, it can be shown that for each consistent random set $\mathcal{S}$, there is a unique, well-justified consonant approximation $\mathcal{S}^*$ whose plausibilistic trace is equal to that of $\mathcal{S}$ [44].

When a consistent random interval is shown as a p-box, it also follows that the trace $\overline{B} - \underline{B} = \pi$ is this same possibility distribution. We have shown [44] that under these conditions, not only is $r = \pi$ a possibility distribution, but moreover is a fuzzy interval as discussed in Section 9.3.3.4. We then call π a **possibilistic histogram**.

An example is shown in Figure 9.13, with m as shown. Note in this case the positive core

$$\mathrm{C} = \bigcap_{A_j \in \mathcal{F}} A_j = [3, 3.5),$$

and that over this region, we have $\overline{B} = 1, \underline{B} = 0, r = 1$.

Returning to the domain of general, finite random sets, the relations among these classes is shown in Figure 9.14. Here we use the term "vacuous" to refer to a random set with a single focal element:

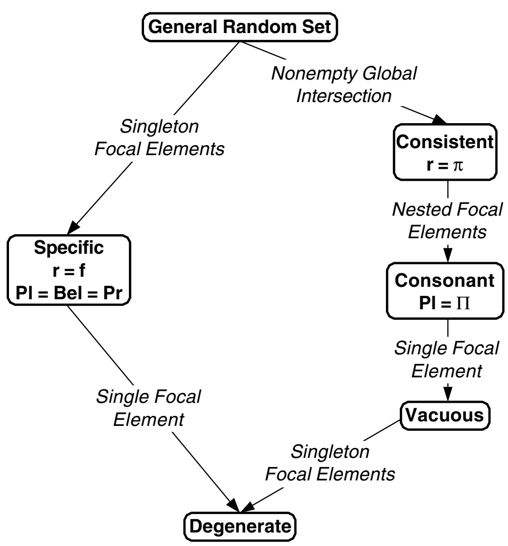

FIGURE 9.14 Relations among classes of random sets.

$\exists A \subseteq \Omega,\ \mathcal{F} = \{A\}$; and "degenerate" to refer to the further case where that single focal element has only one element: $\exists \omega \in \Omega,\ \mathcal{F} = \{\{\omega\}\}$.

9.3.6.3 Interpretations and Applications

Although max-preserving measures have their origins in earlier work, in the context of GIT possibility theory was originally introduced by Zadeh [26] as a kind of information theory strictly related to fuzzy sets. As such, possibility distributions were intended to be measured as and interpreted as linguistic variables.

And, as we have seen, in the context of real-valued fuzzy quantities, it is possible to interpret fuzzy intervals and numbers as possibility distributions. Thus we would hope to use possibility theory as a basis for representing fuzzy arithmetic operations as in Equation 9.8. However, just as there is not a strict symmetry between probabilistic and possibilistic concepts, so there is not a clean generalization here either. In particular, we have shown [45] that the possibilistic properties of fuzzy quantities are not preserved by fuzzy arithmetic convolution operations such as Equation 9.8 outside of their cores and supported.

One of the primary methods for the determination of possibilistic quantities is to take information from a probability distribution and convert it into a possibility distribution. For example, given a discrete probability distribution as a vector $p = \langle p_1, p_2, \ldots, p_n \rangle$, then we can create a possibility distribution $\pi = \langle \pi_1, \pi_2, \ldots, \pi_n \rangle$, where

$$\pi_i = \frac{p_i}{\max p_i}.$$

There are other conversion methods, and an extensive literature, including how the information measure is preserved or not under various transformations [46, 47]. However, one could argue that all such methods are inappropriate: when information is provided in such a way as to be appropriate for a probabilistic approach, then that approach should be used, and *vice versa* [27]. For the purposes of engineering reliability analysis, our belief is that a strong basis for possibilistic interpretations is provided by their grounding in random sets, random intervals, and p-boxes. As we have discussed, this provides a mathematically sound approach for the measurement and interpretation of statistical collections of intervals, which always yield p-boxes, and may or may not yield possibilistic special cases, depending on the circumstances. When they do not, if a possibilistic treatment is still desired, then various normalization procedures are available [48, 49] to transform an inconsistent random set or interval into a consistent or consonant one.

9.3.6.4 Relations between Probabilistic and Possibilistic Concepts

We are now at a point where we can summarize the relations existing between probabilistic and possibilistic concepts in the context of random set theory. To a certain extent, these are complementary, and to another distinct.

Table 9.5 summarizes the relations for general, finite random sets and the special cases of probability and possibility. Columns are shown for the case of general finite random sets, and then the two prominent probabilistic and possibilistic special cases. Note that these describe complete random sets, so that $n = N$ and the indices i on ω_i and j on A_j can be used interchangeably.

In Section 9.3.4 we noted that, formally, all traces of monotone measures are fuzzy sets. Thus, in particular, each probability and possibility distribution is a kind of fuzzy set. In Section 9.3.3.4, we similarly discussed relations among classes of fuzzy quantities, which are fuzzy sets defined on the continuous $\mathbb{R}$. The relations among classes of fuzzy sets defined on a finite space Ω are shown in Figure 9.15, together with examples of each for $\Omega = \{a, b, c, d, e\}$.

Figure 9.16 shows the relations between these distributions or traces as fuzzy sets (quantities defined on the points $\omega \in \Omega$) and the corresponding monotone measures (quantities defined on the subsets $A \subseteq \Omega$). Note that there is not a precise symmetry between probability and possibility. In particular, where probability distributions are symmetric to possibility distributions, probability measures collapse the duality of belief and plausibility, which is exacerbated for possibility.

TABLE 9.5 Summary of Probability and Possibility in the Context of Random Sets

	Random Set	Special Cases: Complete Random Sets, $N = n, i \leftrightarrow j$	
		Probability	Possibility
Focal Sets	Any $A_j \subseteq \Omega$	Singletons: $A_i = \{\omega_i\}$ $\{\omega_i\} = A_i$	Nest: $A_i = \{\omega_1, \ldots, \omega_i\}$ $\{\omega_i\} = A_i - A_{i-1}, A_0 := \varnothing$
Structure	Arbitrary	Finest Partition	Total order
Belief	$\mathrm{Bel}(A) = \sum_{A_j \subseteq A} m_j$	$\Pr(A) := \mathrm{Bel}(A)$	$\eta(A) := \mathrm{Bel}(A)$
Plausibility	$\mathrm{Pl}(A) = \sum_{A_j \cap A \neq \varnothing} m_j$	$\Pr(A) := \mathrm{Pl}(A)$	$\Pi(A) := \mathrm{Pl}(A)$
Relation	$\mathrm{Bel}(A) = 1 - \mathrm{Pl}(A^c)$	$\mathrm{Bel}(A) = \mathrm{Pl}(A) = \Pr(A)$	$\eta(A) = 1 - \Pi(A^c)$
Trace	$\rho_{\mathrm{Pl}}(\omega_i) = \sum_{A_j \ni \omega_i} m_j$	$p(\omega_i) := m(A_j), A_j = \{\omega_i\}$	$\pi(\omega_i) = \sum_{j=i}^{n} m_j$
		$m_i = p_i$	$m_i = \pi_i - \pi_{i+1}$
Measure		$\Pr(A \cup B) = \Pr(A) + \Pr(B)$ $- \Pr(A \cap B)$	$\Pi(A \cup B)$ $= \Pi(A) \vee \Pi(B)$
Normalization		$\sum_i p_i = 1$	$\bigvee_i \pi_i = 1$
Operator		$\Pr(A) = \sum_{\omega_i \in A} p_i$	$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i$
Nonspecificity	$U_{\mathrm{N}}(\mathcal{S}) = \sum_{j=1}^{N} m_j \log_2 \lvert A_j \rvert$		$U_{\mathrm{N}}(\pi) = \sum_{i=2}^{n} \pi_i \log_2 \left[\frac{i}{i-1}\right] =$ $\sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2(i)$
Conflict	$U_{\mathrm{S}}(\mathcal{S}) = -\sum_{i=1}^{n} m(\{\omega_i\}) \log_2(m(\{\omega_i\}))$	$U_{\mathrm{prob}}(p) = -\sum_i p_i \log_2(p_i)$	
Strife	$U_{\mathrm{S}}(\mathcal{S}) = -\sum_{j=1}^{N} m_j \log_2 \left[\sum_{k=1}^{N} m_k \frac{\lvert A_j \cap A_k \rvert}{\lvert A_j \rvert}\right]$	$U_{\mathrm{prob}}(p) = -\sum_i p_i \log_2(p_i)$	
Aggregate Uncertainty	$U_{AU}(\mathcal{S}) = \max_{p: \forall A \subseteq \Omega, \Pr(A) \leq \mathrm{Pl}(A)} U_{\mathrm{prob}}(p)$		

9.4 Conclusion and Summary

This chapter described the basic mathematics of the most common and prominent branches of both classical and generalized information theory. In doing so, we have emphasized primarily a perspective drawing from applications primarily in engineering modeling. These principles have use in many other fields, for example data fusion, image processing, and artificial intelligence.

It must be emphasized that there are many different mathematical approaches to GIT. The specific course of development espoused here, and the relations among the components described, is just one among many. There is a very large literature here that the diligent student or researcher can access.

Moreover, there are a number of mathematical components that properly belong to GIT but space precludes a development of here. In particular, research is ongoing concerning a number of additional mathematical subjects, many of which provide even further generalizations of the already generalized

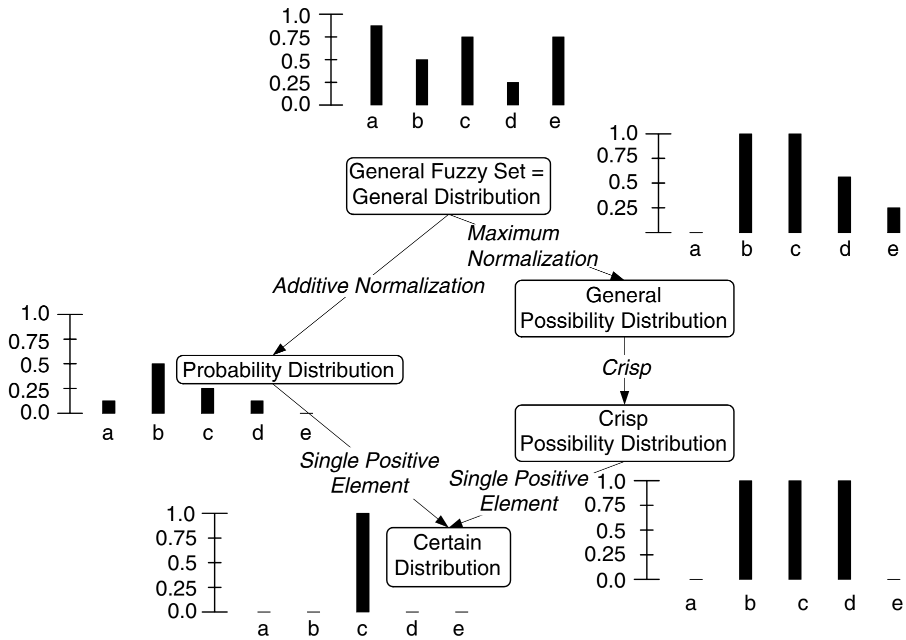

FIGURE 9.15 Relations among classes of general distributions as fuzzy sets.

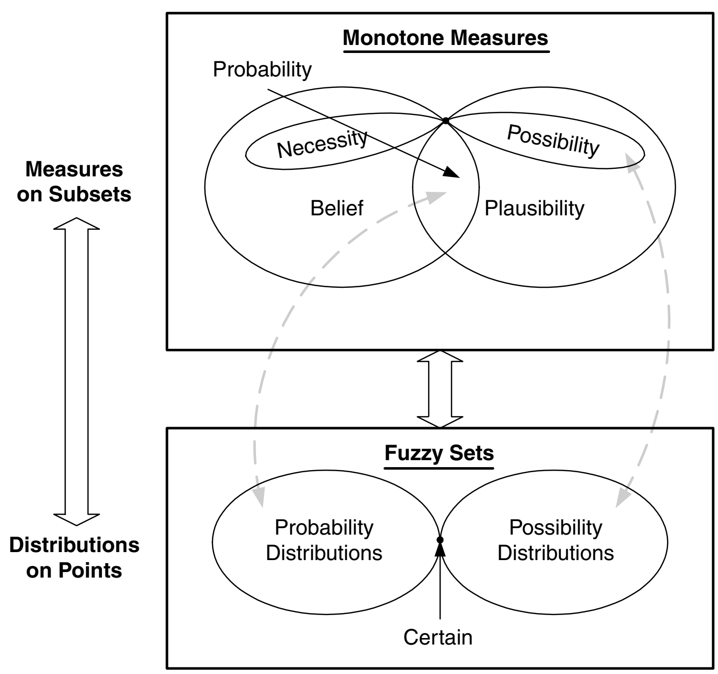

FIGURE 9.16 Relations between distributions on points and measures on subsets.

topics introduced here. These include at least the following:

Rough Sets: Pawlak [50] introduced the structure of a "rough" set, which is yet another way of capturing the uncertainty present in a mathematical system. Let $\mathcal{C} = \{A\} \subseteq 2^{\Omega}$ be a partition of Ω, and assume a special subset $A_0 \subseteq \Omega$ (not necessarily a member of the partition). Then $\mathbf{R}(A_0) := \{\underline{A}_0, \overline{A}_0\}$ is a **rough set** on Ω, where

$$\underline{A}_0 := \{A \in \mathcal{C} : A \subseteq A_0\}, \quad \overline{A}_0 := \{A \in \mathcal{C} : A \cap A_0 \neq \varnothing\}.$$

Rough sets have been closely related to the GIT literature [51], and are useful in a number of applications [52]. For our purposes, it is sufficient to note that $\overline{A}_0$ effectively specifies the support **U**, and $\underline{A}_0$ the nonempty core **C**, of a number of DS structures, and thus an equivalence class of possibility distributions on Ω [53].

Higher-Order Structures: In Section 9.3.3, we described how Zadeh's original move of generalizing from $\{0, 1\}$-valued characteristic functions to $[0, 1]$-valued membership functions can be thought of as a *process* of "fuzzification." Indeed, this lesson has been taken to heart by the community, and a wide variety of fuzzified structures have been introduced. For example, Type II and Level II fuzzy sets arise when fuzzy weights themselves are given weights, or whole fuzzy sets themselves. Or, fuzzified DS theory arises when focal elements *Aj* are generalized to fuzzy subsets $\tilde{A}_j \tilde{\subseteq} \Omega$. There is a fuzzified linear algebra; a fuzzified calculus, etc. It is possible to rationalize these generalizations into systems for generating mathematical structures [54], and approach the whole subject from a higher level of mathematical sophistication, for example by using category theory.

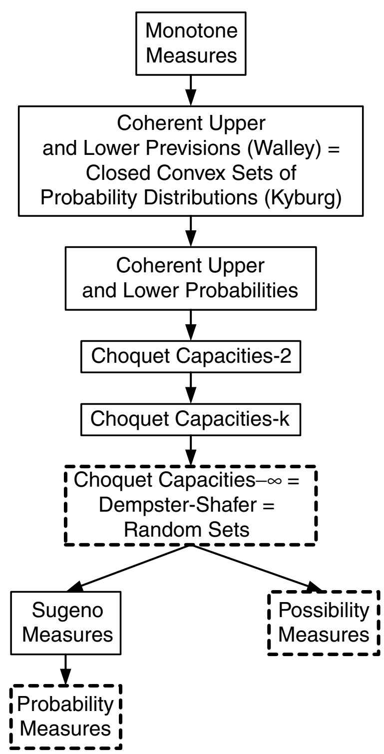

FIGURE 9.17 Relations among classes of monotone measures, imprecise probabilities, and related structures, adapted from [33].

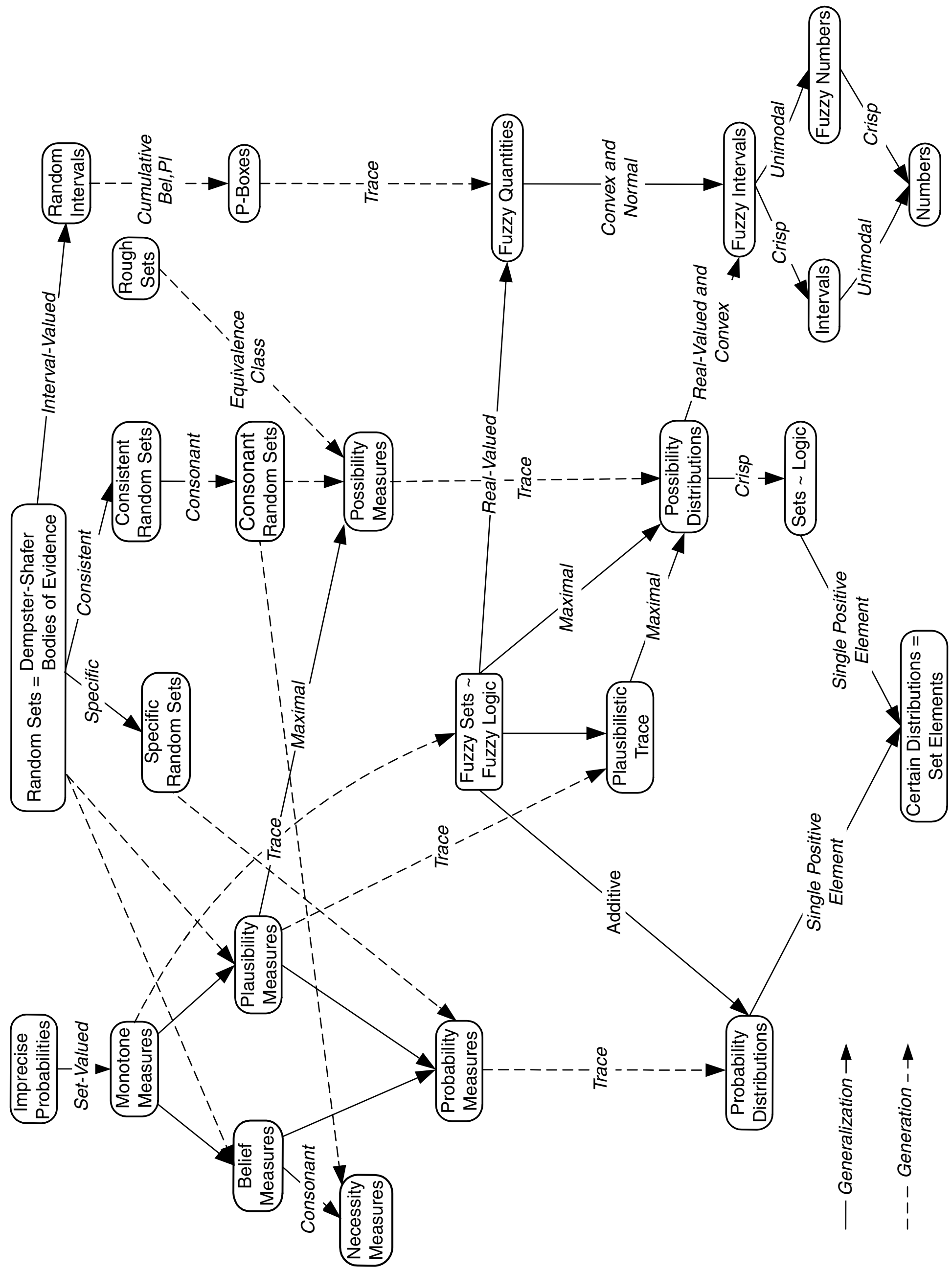

FIGURE 9.18 Map of the various sub-fields of GIT.

Other Monotone Measures, Choquet Capacities, and Imprecise Probabilities: We have focused exclusively on belief and plausibility measures, and their special cases of probability, possibility, and necessity measures. However, there are broad classes of fuzzy or monotone measures outside of these, with various properties worthy of consideration [32]. And while DS theory and random set theory arose in the 1960s and 1970s, their work was presaged by prior work by Choquet in the 1950s that identified many of these classes. Within that context, we have noted above in Section 9.3.1 that belief and plausibility measures stand out as special cases of **infinite order Choquet capacities** [19, 20].

Then, just as monotone measures generalize probability measures by relaxing the additivity property of Equation 9.5, it is also possible to consider relaxing the additivity property of random sets in Equation 9.15, or their range to the unit interval (that is, to consider possibly negative m values), and finally generalizing away from measures on subsets $A \subseteq \Omega$ altogther. Most of the most compelling research ongoing today in the mathematical foundations of GIT concerns these areas, and is described in various ways as **imprecise probabilities** [29] or **convex combinations of probability measures** [33, 55]. Of course, these various generalizations satisfy the extension principle, and thus provide, for example, an alternative basis for the more traditional sub-fields of GIT such as probability theory [56].

Research among all of these more sophisticated classes of generalized measures, and their connections, is active and ongoing. Figure 9.17, adapted from [33], summarizes the current best thought about these relations. Note the appearance in this diagram of certain concepts not explicated here, including previsions, Choquet capacities of finite order, and a class of monotone measures called "Sugeno" measures proper.

We close this chapter with our "grand view" of the relations among most of the structures and classes discussed in Figure 9.18. This diagram is intended to incorporate all of the particular diagrams included earlier in this chapter. Specifically, a solid arrow indicates a mathematical generalization of one theory by another, and thus an instance of where the extension principle should hold. These are labeled with the process by which this specification occurs. A dashed arrow indicates where one kind of structure is generated by another; for example, the trace of a possibility measure yields a possibility distribution.

Acknowledgments

The authors wish to thank the Los Alamos Nuclear Weapons Program for its continued support, especially the Engineering Sciences and Applications Division. This work has also been supported by a research grant from Sandia National Laboratories as part of the Department of Energy Accelerated Strategic Computing Initiative (ASCI).

References

1. Shannon, CE and Weaver, W: (1964) *Mathematical Theory of Communication*, University of Illinois Press, Urbana.
2. Klir, George: (1991) "Generalized Information Theory", *Fuzzy Sets and Systems*, v. **40**, pp. 127–142.
3. Bement, T; Booker, JM; McNulty, S; and N. Singpurwalla: (2003) "Testing the Untestable: Reliaiblity in the 21st Century", *IEEE Transactions in Reliability*, v. **57**, pp. 114–128.
4. Klir, George and Wierman, Mark J: (1997) *Uncertainty-Based Information Elements of Generalized Information Theory,* in: *Lecture Notes in Computer Science*, Creighton University, Omaha.
5. Moore, RM: (1979) *Methods and Applications of Interval Analysis*, in: *SIAM Studies in Applied Mathematics*, SIAM, Philadelphia.
6. Lindley, D: (1987) "Comment: A Tale of Two Wells", *Statistical Science*, v. **2**:1, pp. 38–40.
7. Bayes, T: (1763) "An Essay Towards Solving a Problem in the Doctrine of Chances", *Philosophical Trans. of the Royal Society*, v. **53**, pp. 370–418.

8. Kolmogorov, AN: (1956) *Foundations of the Theory of Probability*, Chelsea, New York.
9. Ben-Haim, Yakov: (2001) *Information-Gap Decision Theory*, Academic Press, London.
10. Sellers, K and Booker, J: (2002) "Bayesian Methods", in: *Fuzzy Logic and Probability Applications: Bridging the*, ed. Ross, Booker, and Parksinson, SIAM, Philadelphia.
11. Graves, T: (2001) "YADAS: An Object-Oriented Framework for Data Analysis Using Markov Chain Monte Carlo", *LANL Technical Report LA-UR-01-4804*.
12. Dubois, Didier and Prade, Henri: (1993) "Fuzzy sets and Probability: Misunderstandings, Bridges and Gaps", in: *Proc. FUZZ-IEEE '93*.
13. Klir, George and Yuan, Bo: (1995) *Fuzzy Sets and Fuzzy Logic*, Prentice-Hall, New York.
14. Klir, George J: (1997) "Uncertainty Theories, Measures, and Principles: An Overview of Personal Views and Contributions", in: *Uncertainty: Models and Measures*, ed. HG Natke and Y Ben-Haim, pp. 27–43, Akademie Verlag, Berlin.
15. Ross, Timothy and Parkinson, J: (2002) "Fuzzy Set Theory, Fuzzy Logic and, Fuzzy Systems", in: *Fuzzy Logic and Probability Applications: Bridging the Gap*, ed. Ross, Booker, and Parksinson, SIAM, Philadelphia.
16. Zadeh, Lotfi A: (1965) "Fuzzy Sets and Systems", in: *Systems Theory*, ed. J. Fox, pp. 29–37, Polytechnic Press, Brooklyn NY.
17. Zadeh, Lotfi A: (1965) "Fuzzy Sets", *Information and Control*, v. **8**, pp. 338–353.
18. Dempster, AP: (1967) "Upper and Lower Probabilities Induced by a Multivalued Mapping", *Annals of Mathematical Statistics*, v. **38**, pp. 325–339.
19. Lamata, MT and Moral, S: (1989) "Classification of Fuzzy Measures", *Fuzzy Sets and Systems*, v. **33**, pp. 243–153.
20. Sims, John R and Wang, Zhenyuan: (1990) "Fuzzy Measures and Fuzzy Integrals: An Overview", *Int. J. of General Systems*, v. **17**:2–3, pp. 157–189.
21. Shafer, Glen: (1976) *Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.
22. Kendall, DG: (1974) "Foundations of a Theory of Random Sets", in: *Stochastic Geometry*, ed. EF Harding and DG Kendall, pp. 322–376, Wiley, New York.
23. Matheron, G: (1975) *Random Sets and Integral Geometry*, Wiley, New York.
24. Dempster, AP: (1968) "Upper and Lower Probabilities Generated by a Random Interval", *Annals of Mathematical Statistics*, v. **39**:3, pp. 957–966.
25. Sugeno, Michio: (1972) "Fuzzy Measures and Fuzzy Integrals", *Trans. SICE*, v. **8**:2.
26. Zadeh, Lotfi A: (1978) "Fuzzy Sets as the Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, v. **1**, pp. 3–28.
27. Joslyn, Cliff: (1995) "In Support of an Independent Possibility Theory", in: *Foundations and Applications of Possibility Theory*, ed. G de Cooman et al., pp. 152–164, World Scientific, Singapore.
28. Klir, George and Folger, Tina: (1987) *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall.
29. Walley, P: (1990) *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, New York.
30. Singpurwalla, N and Booker, J: (2002) "Membership Functions and Probability Measures of Fuzzy Sets", *Los Alamos Technical Report LAUR 02-0032*., LANL.
31. Booker, JM and Singpurwalla, N: (2003) "Using Probability and Fuzzy Set Theory for Reliability Assessment", in: *Proc. 21st Int. Modal Analysis Conference (IMAC-XXI)*, pp. 259, Kissimmee FL.
32. Wang, Zhenyuan and Klir, George J: (1992) *Fuzzy Measure Theory*, Plenum Press, New York.
33. Klir, George and Smith, Richard M: (2000) "On Measuring Uncertainty and Uncertainty-Based Information: Recent Developments", *Annals of Mathematics and Artificial Intelligence*, v. **32**:1-4, pp. 5–33.
34. Sentz, Kari and Ferson, Scott: (2002) "Combination Rules in Dempster-Shafer Theory", in: *6th World Multi-Conference on Systemic, Cybernetics, and Informatics.*
35. Joslyn, Cliff: (1996) "Aggregation and Completion of Random Sets with Distributional Fuzzy Measures", *Int. J. of Uncertainty, Fuzziness, and Knowledge-Based Systems*, v. **4**:4, pp. 307–329.
36. Dubois, Didier and Prade, Henri: (1990) "Consonant Approximations of Belief Functions", *Int. J. Approximate Reasoning*, v. **4**, pp. 419–449.

37. Helton, JC and Oberkampf, WL, eds.: (2004) "Special Issue: Alternative Representations of Epistemic Uncertainty", *Reliability Engineering and Systems Safety*, 95:1-3.
38. Joslyn, Cliff and Helton, Jon C: (2002) "Bounds on Plausibility and Belief of Functionally Propagated Random Sets", in: *Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2002)*, pp. 412–417.
39. Joslyn, Cliff and Kreinovich, Vladik: (2002) "Convergence Properties of an Interval Probabilistic Approach to System Reliability Estimation", *Int. J. General Systems,* in press.
40. Oberkampf, WL; Helton, JC; Wojtkiewicz, SF, Cliff Joslyn, and Scott Ferson: (2004) "Uncertainty in System Response Given Uncertain Parameters", *Reliability Reliability Engineering and System Safety*, 95:1-3, pp. 11–20.
41. Tonon, Fulvio and Bernardini, Alberto: (1998) "A Random Set Approach to the Optimization of Uncertain Structures", *Computers and Structures*, v. **68**, pp. 583–600.
42. Ferson, Scott; Kreinovich, V; and Ginzburg, L et al.: (2002) "Constructing Probability Boxes and Dempster-Shafer Structures", *SAND Report 2002-4015*, Sandia National Lab, Albuquerque NM, http://www.sandia.gov/epistemic/Reports/SAND2002-4015. pdf.
43. Yager, Ronald R: (1986) "Arithmetic and Other Operations on Dempster-Shafer Structures", *Int. J. Man-Machine Studies*, v. **25**, pp. 357–366.
44. Joslyn, Cliff: (1997) "Measurement of Possibilistic Histograms from Interval Data", *Int. J. General Systems*, v. **26**:1-2, pp. 9–33.
45. Joslyn, Cliff and Ferson, Scott: (2003) "Convolutions of Representations of Random Intervals", in preparation.
46. Joslyn, Cliff: (1997) "Distributional Representations of Random Interval Measurements", in: *Uncertainty Analysis in Engineering and the Sciences*, ed. Bilal Ayyub and Madan Gupta, pp. 37–52, Kluwer.
47. Sudkamp, Thomas: (1992) "On Probability-Possibility Transformations", *Fuzzy Sets and Systems*, v. **51**, pp. 73–81.
48. Joslyn, Cliff: (1997) "Possibilistic Normalization of Inconsistent Random Intervals", *Advances in Systems Science and Applications*, special issue, ed. Wansheng Tang, pp. 44–51, San Marcos, TX, ftp://wwwc3.lanl.gov/pub/users/joslyn/iigss97.pdf.
49. Oussalah, Mourad: (2002) "On the Normalization of Subnormal Possibility Distributions: New Investigations", *Int. J. General Systems*, v. **31**:3, pp. 227–301.
50. Pawlak, Zdzislaw: (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Boston.
51. Dubois, Didier and Prade, Henri: (1990) "Rough Fuzzy Sets and Fuzzy Rough Sets", *Int. J. General Systems*, v. **17**, pp. 191–209.
52. Lin, TY and Cercone, N, eds.: (1997) *Rough Sets and Data Mining: Analysis of Imprecise Data*, Kluwer, Boston.
53. Joslyn, Cliff: (2003) "Multi-Interval Elicitation of Random Intervals for Engineering Reliability Analysis", *Proc. 2003 Int. Symp. on Uncertainty Modeling and Analysis (ISUMA 03)*.
54. Joslyn, Cliff and Rocha, Luis: (1998) "Towards a Formal Taxonomy of Hybrid Uncertainty Representations", *Information Sciences*, v. **110**:3-4, pp. 255–277.
55. Kyburg, HE (1987): "Bayesian and Non-Bayesian Evidential Updating", *Artificial Intelligence*, v. **31**, pp. 271–293.
56. de Cooman, Gert and Aeyels, Dirk: (2000) "A Random Set Description of a Possibility Measure and Its Natural Extension", *IEEE Trans. on Systems, Man and Cybernetics A*, v. **30**:2, pp. 124–130.