Possibilistic Systems Within a General Information Theory

Dr. Cliff Joslyn
Distributed Knowledge Systems Team
Computer and Information Research Group (CIC-3)
Los Alamos National Laboratory*

Abstract

We survey possibilistic systems theory and place it in the context of Imprecise Probabilities and General Information Theory (GIT). In particular, we argue that possibilistic systems hold a distinct position within a broadly conceived, synthetic GIT. Our focus is on systems and applications which are semantically grounded by empirical measurement methods (statistical counting), rather than epistemic or subjective knowledge elicitation or assessment methods. Regarding fuzzy measures as special previsions, and evidence measures (belief and plausibility measures) as special fuzzy measures, thereby we can measure imprecise probabilities directly and empirically from set-valued frequencies (random set measurement). More specifically, measurements of random intervals yield empirical fuzzy intervals. In the random set (Dempster-Shafer) context, probability and possibility measures stand as special plausibility measures in that their "distributionality" (decomposability) maps directly to an "aggregable" structure of the focal classes of their random sets. Further, possibility measures share with imprecise probabilities the ability to better handle "open world" problems where the universe of discourse is not specified in advance. In addition to empirically grounded measurement methods, possibility theory also provides another crucial component of a full systems theory, namely prediction methods in the form of finite (Markov) processes which are also strictly analogous to the probabilistic forms.

Keywords. Possibility Theory, random sets, fuzzy measures, imprecise probabilities, general information theory, possibilistic processes.

1 Possibility Theory and Imprecise Probabilities in General Information Theory

A central conern for interdisciplinary scientists is the search for properties which can be measured across systems of different types: if we assert that two systems actually have the same structure or organization, how can that hypothesis become well-posed and testable? Such questions are usually framed in a relational language of such concepts as order, organization, structure, variety, constraint, freedom, determinism, and complexity. A formal theory of relational concepts has rested classically on information theories, and in particular on concepts of information, such as Shannon's statistical entropy, which are defined as a reduction in or lack of uncertainty. In turn, these uncertainty-based information theories were rooted deeply within the formalism of traditional probability theory, with a corresponding emphasis on entropy measures, Monte Carlo methods, Bayes nets, Markov models, etc.

This view is currently expanding in two significant ways. First, there has been progress towards addressing a primary criticism of information theory, namely that it is purely *syntactic* and does not involve anything about the *meaning* of the signal. There is thus a growing **semiotic** theory of information, where issues of the semantics, interpretation of signals, and the groundings of signals in measurements are finally being seriously considered [27].

Second, since the introduction of fuzzy sets [26] and evidence theory [4, 29] in the mid-1960's there has been a proliferation of mathematical methods for the representation of uncertainty which generalize beyond classical probability theory [25]. In addition to a fully developed fuzzy systems theory [26], there are also fuzzy measures [33], rough sets [28], random sets [8, 22] (Dempster-Shafer bodies of evidence [9, 29]), and possibilistic systems [2]. There is a pressing need to

*Mail Stop B265, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, joslyn@lanl.gov, www.c3.lanl.gov/~joslyn, (505) 667-9096.

synthesize these fields within a collective as General Information Theory (GIT) [24], searching out larger formal frameworks within which to place these various components with respect to each other. And indeed there is a growing movement in that direction [6, 21, 25].

In particular, Imprecise Probabilities have been advanced as providing a grand generalization of all of these methods [3, 30]. As a general framework, imprecise probabilities have both advantages and disadvantages for particular interpretations and applications. On the one hand, they can subsume multiple other representations. On the other hand, in their general form they are complex mathematical structures, whose primary interpretations and measurements are grounded in epistemic evaluations. More specialized mathematical frameworks, for example within fuzzy measures or probability or possibility theory, are more constrained structures, with the tradeoff of potentially greater applicability against less generality.

Our work specifically is motivated by the introduction of possibility theory as the first alternative, non-probabilistic form of information theory [23], and thus as a branch of GIT [2]. Within GIT, possibility theory is unique in that it provides structures and methods which parallel traditional information theory, with strict possibilistic correlates to distributions, entropy measures, Markov processes, and Monte-Carlo methods, etc. [11, 12, 14]. Simultaneously, through random-set based measurements, empirical methods are available for measurement of possibilistic structures, including histograms and sample statistics.

Furthermore, there is evidence to support the claim that these possibilistic forms are unique in providing such a close parallel to the standard probabilistic forms. The understanding of the deep connections between possibility measures, coherent upper previsions, and random sets [1], and the fact that like imprecise probabilities, possibilistic systems are better able to handle "open-world" problems with unspecified or changing universes of discourse [32], combine to suggest the way forward to integrating possibilistic systems theory within the broad context of a GIT involving imprecise probabilities.

In the rest of this paper we briefly survey aspects of possibilistic systems theory and place them in the context of imprecise probability and GIT. In particular, we recognize possibility measures as extreme plausibility measures, which in turn are fuzzy measures, and finally which in turn can be cast as previsions on sets. In this way possibilistic systems are available in an imprecise probability context.

We consider in particular three aspects of possibilistic systems theory:

- We can measure imprecise probabilities directly and empirically from set-valued frequencies (random set measurement), and derive empirical fuzzy numbers and intervals from random interval measurements.
- Given a semantic grounding in random set (Dempster-Shafer) measurement, we then understand that those which yield probability and possibility measures are special in that they are t-conorm distributional (decomposable) and also have certain simple topologies.
- Finally, in addition to empirically grounded measurement methods, possibility theory also provides another crucial component of a full systems theory, namely prediction methods in the form of finite (Markov) processes which are also strictly analogous to their probabilistic form.

2 Random Set Approach to Possibility Theory

Assume a universe of discourse $\Omega = \{\omega\}$. We generally consider $\Omega = \{\omega_i\}, 1 \leq i \leq n$ to be finite, although sometimes we will recognize that $\Omega = \mathbb{R}$, and consider half-open interval subsets, elements of the class denoted $\mathcal{D} := \{[a,b) \subseteq \mathbb{R} : a,b \in \mathbb{R}, a < b\}$. Given a class $\mathcal{C} = \{A\} \subseteq 2^\Omega$, define the core as $\mathbf{C}(\mathcal{C}) := \bigcap_{A \in \mathcal{C}} A$.

Define a **triangular conorm** $\sqcup\colon [0,1]^2 \mapsto [0,1]$ (resp. **triangular norm** $\sqcap\colon [0,1]^2 \mapsto [0,1]$) as an associative, commutative, monotonic operator with identity 0 (resp. 1). $\mathcal{R} := \langle \sqcup, \sqcap \rangle$ is a **conorm semiring** if $\sqcap$ distributes over $\sqcup$.

The function $\nu\colon 2^\Omega \mapsto [0,1]$ is a (finite) fuzzy measure [33] if $\nu(\emptyset) = 0$ and $\forall A, B \subseteq \Omega, A \subseteq B \rightarrow \nu(A) \leq \nu(B)$. ν is called **distributional** if there exists a conorm $\sqcup$ such that $\forall A \subseteq \Omega, \bigsqcup_{\omega_i \in A} q_\nu(\omega_i) = \nu(A)$, where $q_\nu\colon \Omega \mapsto [0,1]$, with $q_\nu(\omega_i) := \nu(\{\omega_i\})$ the **distribution** of ν. Furthermore, ν is normal when $\nu(\Omega) = 1$, so that $\bigsqcup_{\omega_i \in \Omega} q_\nu(\omega_i) = 1$. For a fixed finite fuzzy measure ν, denote $\vec{q} = \langle q_i \rangle := \langle q_\nu(\{\omega_i\}) \rangle$ for $1 \leq i \leq n$.

Consider a a probability measure Pr with probability distribution $p := q_{\Pr}, \vec{p} = \langle p_i \rangle := \vec{q}$ which is an additively normal fuzzy measure with $\sum_{i=1}^n p_i = 1$. Then Pr is a $+_b$-distributional fuzzy measure where $x +_b y := (x+b) \wedge 1$, $x, y \in [0,1]$ and $\wedge$ is the minimum operator. The standard forms of probability result $\forall A, B \subseteq \Omega$:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B),$$

$$\Pr(A) = \sum_{\omega_i \in A} p_i, \qquad \sum_{i=1}^{n} p_i = 1.$$

The central tenet of possibility theory is the introduction of a fuzzy measure Π with possibility distribution $\pi := q_\Pi, \vec{\pi} = \langle \pi_i \rangle := \vec{q}$ which is distributional for $\sqcup = \vee$, the maximum operator. The equations of probability now take the form $\forall A, B \subseteq \Omega$

$$\Pi(A \cup B) = \Pi(A) \vee \Pi(B),$$

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i, \qquad \bigvee_{i=1}^{n} \pi_i = 1. \qquad (1)$$

Possibility measures and distributions share a great advantage with imprecise probabilities, at least as advanced by Walley in the imprecise Dirichlet approach [32], in that they can adequately represent "open-world" problems where the universe of discourse is either unspecified or changes. In particular, given a possibility distribution on Ω, if Ω is updated, then no global recalculation of π is required. This is because the maximal possibilistic normalization is a "local" property of the core $\mathbf{C}(\pi) := \{\omega : \pi(\omega) = 1\} \subseteq \Omega$, and not a global property of the whole distribution, as with additive probability distributions.

Possibility measures are usually intepreted in the context of fuzzy sets, and in particular the possibility distribution π is interpreted as a fuzzy set. Their measurements are then grounded in traditional fuzzy systems methods of subjective evaluations [31]. An alternative approach is to ground the measurement of possibility measures and distributions in the context of empirically-derived random sets.

Given a probability space $\langle X, \Sigma, \Pr \rangle$, then a function $S: X \mapsto 2^\Omega - \{\emptyset\}$, where $-$ is set subtraction, is a random subset of Ω if S is Pr-measurable, so that $\forall \emptyset \neq A \subseteq \Omega, S^{-1}(A) \in \Sigma$. In the finite case, they can be seen more simply as random variables taking values on subsets of Ω. Further, they are mathematically isomorphic to bodies of evidence in Dempster-Shafer evidence theory [4, 29]. In this context, we can introduce a function $m: 2^\Omega \mapsto [0,1]$ as an **evidence function** (basic assignment) when $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. Then $\mathcal{S} := \{\langle A_j, m_j \rangle : m_j > 0\}$ is a finite random set where $A_j \subseteq \Omega, m_j := m(A_j)$, and $1 \leq j \leq N := |\mathcal{S}| \leq 2^n - 1$. Denote the focal set of $\mathcal{S}$ as the class $\mathcal{F}(\mathcal{S}) := \{A_j : m_j > 0\} \subseteq 2^\Omega$. A random set $\mathcal{S}$ is **consistent** if $\mathbf{C}(\mathcal{F}(\mathcal{S})) \neq \emptyset$.

The plausibility and belief measures on $\forall A \subseteq \Omega$ are

$$\mathrm{Pl}(A) := \sum_{A_j \not\perp A} m_j, \qquad \mathrm{Bel}(A) := \sum_{A_j \subseteq A} m_j,$$

where $A \perp B$ denotes $A \cap B = \emptyset$. The plausibility assignment (otherwise known as the **trace** or one-point coverage) of $\mathcal{S}$ is $\vec{\rho}(\mathcal{S}) = \langle \rho_i \rangle$, where $\rho_i := \mathrm{Pl}(\{\omega_i\}) = \sum_{A_j \ni \omega_i} m_j$. Clearly ρ is a fuzzy set.

de Cooman and Aeyels have provided full details on the imprecise probability interpretation of possibility measures and random sets [1]. Note that [1, 30, 33]:

- Pl and Bel are generally non-additive fuzzy measures without distributions, and are dual, in that $\forall A \subseteq \Omega, \mathrm{Bel}(A) = 1 - \mathrm{Pl}(\overline{A})$.
- Π is an extreme plausibility, whose dual belief is the necessity measure $\eta(A) := 1 - \Pi(\overline{A})$.
- Pr is both a plausibility and its self-dual belief measure.
- If Pl and Bel are normal, then they are coherent upper and lower probabilities on the events $A \subseteq \Omega$.
- Π is normal iff it is a coherent upper prevision on the events.

Given a random set $\mathcal{S}$, if Pl has a distribution operator $\sqcup$, then $\vec{q}(\mathcal{S}) := \vec{\rho}(\mathcal{S})$ is called the distribution of $\mathcal{S}$. In particular, when

$$\forall A_j \in \mathcal{F}(\mathcal{S}), \qquad |A_j| = 1, \qquad (2)$$

then $\mathcal{S}$ is called specific, $\Pr(A) := \mathrm{Pl}(A) = \mathrm{Bel}(A)$ becomes a probability measure, and $\vec{p}(\mathcal{S}) := \vec{q}(\mathcal{S}) = \vec{\rho}(\mathcal{S})$ is a probability distribution. Similarly, $\mathcal{S}$ is called consonant ($\mathcal{F}(\mathcal{S})$ is a nest) when (without loss of generality for ordering, and letting $A_0 := \emptyset$) $A_{j-1} \subseteq A_j$. Now $\Pi(A) := \mathrm{Pl}(A)$ is a possibility measure and $\eta(A) := \mathrm{Bel}(A)$ is a necessity measure.[1] $\vec{\pi} := \vec{q}(\mathcal{S}) = \vec{\rho}(\mathcal{S})$ is then a possibility distribution.

Each random set $\mathcal{S}$ maps to a unique fuzzy set $\vec{\rho}(\mathcal{S})$, or to its distribution $\vec{q}(\mathcal{S})$ if $\sqcup$ exists. But when we begin with a particular fuzzy set $\mu: \Omega \mapsto [0,1]$, or in vector form $\vec{\mu}$, there is generally a non-empty, non-unique equivalence class of random sets $\Psi(\vec{\mu})$ for which $\forall \mathcal{S} \in \Psi(\vec{\mu}), \vec{\rho}(\mathcal{S}) = \vec{\mu}$ [7]. When $\vec{\mu}$ begins as an additive probability distribution $\vec{p}$, then $|\Psi(\vec{p})| = 1$, so that $\vec{p}$ uniquely determines a specific (in the sense of (2)) random set.

[1] Since results for necessity are dual to those of possibility, only possibility will be discussed in the sequel.

But when $\vec{\mu}$ begins as a maximal possibility distribution $\vec{\pi}$, then in general $|\Psi(\vec{\pi})| > 1$. All of the $\mathcal{S} \in \Psi(\vec{\pi})$ are consistent, and thus it is this consistency which is both necessary and sufficient for $\mathcal{S}$ to have a maximally normalized possibility distribution $\vec{\pi} = \vec{\rho}(\mathcal{S})$ by (1). In particular, $\mathcal{S}$ is consistent iff $\bigvee_{i=1}^n \rho_i = 1$. Then while $\mathcal{S}$ might not be consonant and Pl not a possibility measure, there is a unique approximating possibility measure Π^* and consonant random set $\mathcal{S}^*(\vec{\pi}(\mathcal{S})) \in \Psi(\vec{\pi})$. Thus in general when working with possibility theory in the context of finite random sets, a consistent random set $\mathcal{S}$ is a sufficient condition to generate a possibility distribution $\vec{\pi}(\mathcal{S})$.

As we consider possibilistic measurement proper below, it will be desirable to let $\Omega = \mathbb{R}$. A random interval, denoted $\mathcal{A}$, is 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 $\mathbb{R}$. The trace of $\mathcal{A}$ is then $\rho_{\mathcal{A}}: \mathbb{R} \mapsto [0,1]$, where $\forall x \in \mathbb{R}, \rho_{\mathcal{A}}(x) := \mathrm{Pl}(\{x\}) = \sum_{A_j \ni x} m_j$. A fuzzy subset of the real line $\widetilde{F} \,\widetilde{\subseteq}\, \mathbb{R}$ is a **fuzzy interval** if $\widetilde{F}$ is maximally normalized and convex, so that $\forall x, y \in \mathbb{R}, \forall z \in [x,y], \mu_{\widetilde{F}}(z) \geq \mu_{\widetilde{F}}(x) \wedge \mu_{\widetilde{F}}(y)$. A **fuzzy number** is a fuzzy interval $\widetilde{F}$ where $\exists x \in \mathbb{R}, \mathbf{C}(\widetilde{F}) = \{x\}$.

3 Measurement of Possibilistic Histograms

Random set counting provides a superb empirical method to ground the measurent of evidence (belief and plausibility) measures. They are a direct generalization of traditional frequentist methods to measure probability distributions (for an empirical approach using imprecise probabilities, see [32]). More specifically, measurement of consistent random intervals yield empirical fuzzy intervals. Full formal details of the following are available elsewhere [17, 19].

3.1 Random Set Measurement

The central concept is the introduction of a **general measuring device**, a system $\mathcal{M} := \left\langle \mathcal{C}, \vec{B}, C \right\rangle$, where:

- $\mathcal{C} := \{A_{j'}\} \subseteq 2^\Omega, A_{j'} \neq \emptyset, 1 \leq j' \leq N'$ is the class of **observable sets**;
- $\vec{B} := \langle B_s \rangle$ is the **general measurement record**, a vector of each observed subset for $1 \leq s \leq M$, so that $\forall B_s \in \vec{B}, \exists ! A_{j'} \in \mathcal{C}, B_s = A_{j'}$; and
- $C: \mathcal{C} \mapsto \mathcal{W}$ is the **set counting function**, where $\forall A_{j'} \in \mathcal{C}, C_{j'} := C(A_{j'})$ is the number of occurrences of $A_{j'}$ in $\vec{B}$.

The nature of the measuring device will depend on the elements and topological structure of $\mathcal{C}$. Generally, $\mathcal{C}$ is an arbitrary collection of possibly non-disjoint subsets. In a classical measuring device like a thermometer, $\mathcal{C} = \{B_s\}$ would be a collection of disjoint, equal length, half-open intervals $B_s = [d_s, d_{s+1})$.

Given a general measuring device $\mathcal{M}$, let $\mathcal{F}^E := \{A_j\} \subseteq \mathcal{C}, 1 \leq j \leq N$, be an **empirical focal set** derived by eliminating the duplicates from $\vec{B}$, where:

$$1 \leq j \leq N, \qquad \mathcal{F}^E \subseteq \mathcal{C}, \qquad N \leq N',$$

$$N \leq M, \qquad \forall A_j \in \mathcal{F}^E, \exists B_s \in \vec{B}, A_j = B_s.$$

$\mathcal{F}^E$ is essentially the restriction of $\mathcal{C}$ to those subsets which are actually observed in the record $\vec{B}$.

Finally, we can construct the **set-frequency distribution** function $m^E: \mathcal{F}^E \mapsto [0,1]$ where

$$m^E(A_j) := \frac{C_j}{\sum_{A_j \in \mathcal{F}^E} C_j} = \frac{C_j}{M}, \qquad m_j^E := m^E(A_j).$$

It follows that m^E is an evidence function, and thus it is possible to derive an empirical random set $\mathcal{S}^E$ whose focal set is $\mathcal{F}^E$.

The topological properties of $\mathcal{C}$ (partially) flow to $\mathcal{S}^E$. In particular, if $\mathcal{C}$ is point-valued, in that $\mathcal{C} = \{\{\omega_{j'}\}\}$ for some collection of $\omega_{j'} \in \Omega$, then of course the trace $\vec{\rho}$ is just a probability distribution. More generally, if $\mathcal{M}$ is a classical device like the thermomemter, then the A_j are disjoint, and this degenerates to traditional relative frequencies, considering the A_j as points in some simplified meta-space. But it also might be that $\mathcal{S}^E$ is consonant or consistent, yielding $\vec{\rho}$ a possibility distribution.

3.2 Possibilistic Histograms from Random Intervals

An important random interval case is when $\Omega = \mathbb{R}$ and $\mathcal{C} \subseteq \mathcal{D}$. Then the empirical random set $\mathcal{S}^E$ becomes an empirical random interval $\mathcal{A}^E$ with plausibilistic trace $\rho_{\mathcal{A}}$. If $\mathcal{A}^E$ is consistent then $\pi^E := \rho_{\mathcal{A}^E}$ is called a **possibilistic histogram**, which is an empirically-derived possibility distribution. If $\mathcal{A}^E$ is not consistent, then various possibilistic approximations are available, in particular interval versions of focused consistent transformations [20, 18].

The approach is illustrated in Fig. 1. In the example, $\mathcal{C} = \mathcal{D}$, and $M = 4$ intervals are observed in $\vec{B}$. $\mathcal{F}^E$ and m are shown, with $N = 3$. Note that $\mathcal{F}^E$ is consistent, and thus the derived π^E is shown on the right.

If $\mathcal{F}^E$ is consistent, it follows that not only is π^E a possibility distribution, it is also a fuzzy interval.

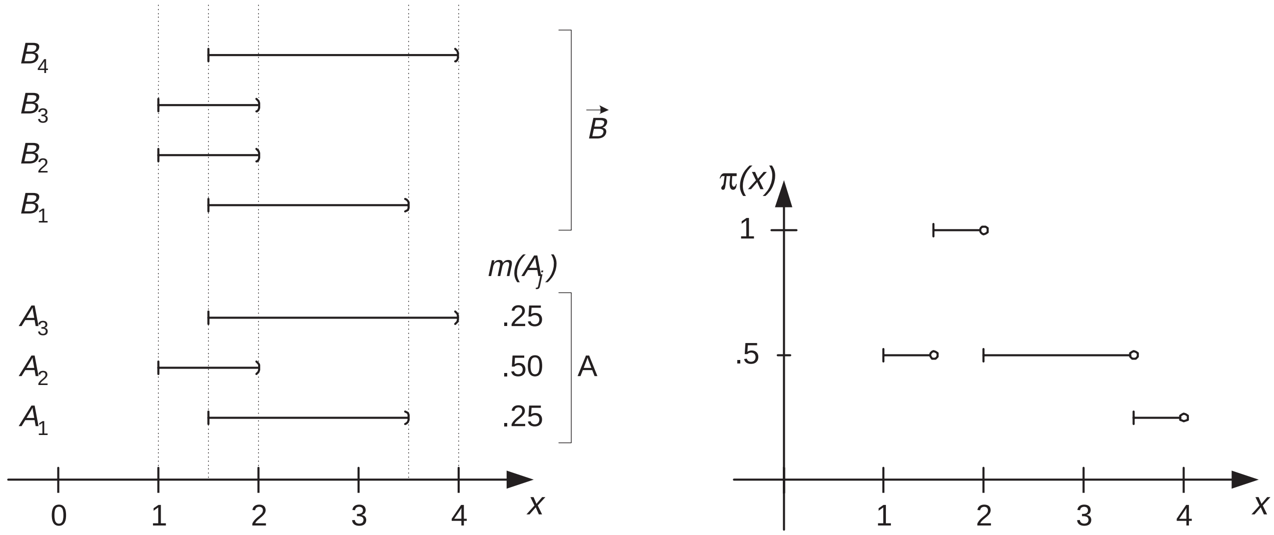

Figure 1: (Left) Observed focal elements of a random interval. (Right) Possibilistic histogram π^E.

These are classical structures in fuzzy theory and used in a variety of applications. But in this "raw" form the π^E are piecewise constant, with at least $N+1$ and at most $2N$ discontinuities. Various methods are available to derive continuous fuzzy interval and fuzzy number forms, yielding traditional forms which still preserve most of the characteristics of the "raw" histograms [17].

A simple example is shown in Fig. 2. On the left, below are two observed intervals which yield a consonant empirical random interval. Above is the corresponding "raw" possibilistic histogram. On the right the histogram is repeated, and also shown are three candidate continuous approximations, including a triangular fuzzy number and a trapezoidal fuzzy interval. The inner continuous approximation is also a possibility, amongst others.

4 Distributional and Aggregable Random Sets

While we ground possibilistic measurement in a random set (interval) context, in general random intervals are not consistent, and yield plausibility measures which are not distributional. We strive to produce distributional possibility measures since they have the great advantage of being able to be constructed on the basis of on the order of n point values, rather than $N \approx 2^n$ set values. In the random set context, probability and possibility measures stand out as special, in that this distributionality is paired with the simple topological structure of the random set. Full formal details of the following are available elsewhere [15].

A random set $\mathcal{S}$ is called g-aggregable is there is a one to one function $g \colon \mathcal{F}(\mathcal{S}) \mapsto \Omega$ called a **structural aggregation function**. One might think to require $g(A_j) \in A_j$, but the results are equivalent to a permutation of Ω.

If $\mathcal{S}$ is g-aggregable, then denote the **numerical aggregation function** $h \colon \mathcal{S} \mapsto [0,1]$ with $h(m_j) = \mathrm{Pl}(g(A_j))$. g maps each focal element A_j to a universe element $g(A_j)$, and h maps that to its plausibility assignment value $h(m_j)$. In general, a random set $\mathcal{S}$ may have multiple g corresponding to the various permutations of the A_j and ω_i.

A random set $\mathcal{S}$ is g-aggregable iff $|\mathcal{S}| = N \leq |\Omega| = n$. If this becomes equality, then $\mathcal{S}$ is called g-complete. If a g-complete random set $\mathcal{S}$ is also $\sqcup$-distributional, then the distribution $\vec{\mathrm{Pl}}$ is called complete. In a g-complete random set, the focal elements and universe elements are mutually determining, with each focal element A_j existing as a particular $g^{-1}(\omega_j)$. The indices i and j are then identical and can be used interchangeably. Also then g is onto, with inverse $g^{-1}(\omega_j) = A_j$, and h^{-1} may also exist, so that $m_j = h^{-1}(\mathrm{Pl}_j)$.

Random sets yielding probability and possibility measures as their plausibility measures are special in that they are both distributional and aggregable. It remains to be proved that they are unique in this respect, but the evidence is highly suggestive. In particular [15]:

- Probability is characterized by disjointness of the random set and the additivity of plausibility.

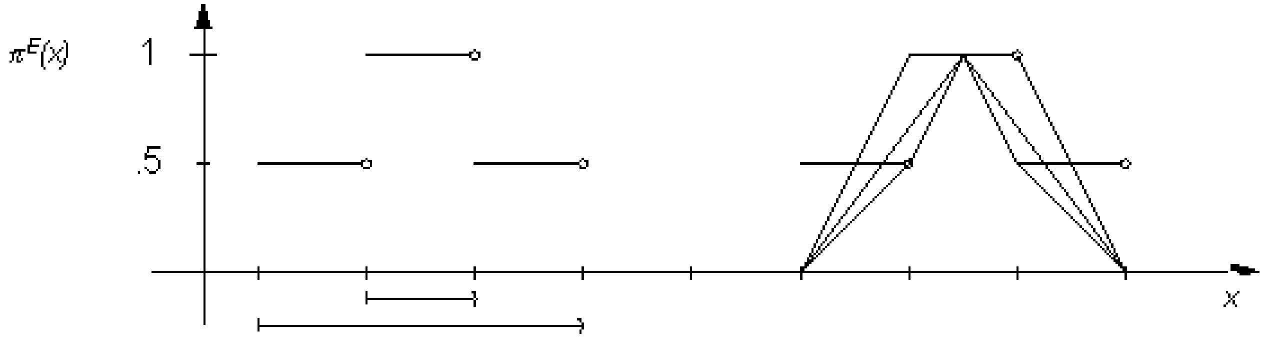

Figure 2: (Left) Below: Two observed intervals. Above: Piecewise constant possibilistic histogram. (Right) Three example piecewise linear continuous approximations.

When g is defined such that $\forall A_i, \exists! \omega_j, g(A_i) = \omega_j$, then $\mathcal{S}$ is specific. Then Pl = Pr, which is also $+_b$-distributional. If $\mathcal{S}$ is also complete, then

$$g(A_j) = \omega_j, \qquad g^{-1}(\omega_j) = A_j,$$

$$h(m_j) = h^{-1}(p_j) = p_j = m_j,$$

and $\forall \omega_i \in \Omega, p_i > 0$.

- Possibility is characterized by nestedness of the random set and the maximality of plausibility. When g is defined such that $\forall A_j, g(A_j) \in A_j - A_{j-1}$, then $\mathcal{S}$ is called consonant. Then Pl = Π, which is also $\vee$-distributional. If $\mathcal{S}$ is also complete, then

$$g(A_j) = A_j - a_{j+1} = \omega_j$$

$$g^{-1}(\omega_j) = \{\omega_1, \omega_2, \ldots, \omega_j\} = A_j,$$

$$h(m_j) = \sum_{k=j}^{N} m_k = \pi_j,$$

$$h^{-1}(\pi_j) = \pi_j - \pi_{j+1} = m_j,$$

$$1 = \pi_1 > \pi_2 > \cdots > \pi_n > 0.$$

- In search of other information theories, we first consider Sugeno-distributional fuzzy measures [5, 15] defined by

$$\nu_\lambda(A \cup B) := \nu_\lambda(A) + \nu_\lambda(B) + \lambda \nu_\lambda(A) \nu_\lambda(B),$$

$$A \not\subseteq B, \quad \lambda \in (-1, \infty).$$

ν_λ is $\sqcup_\lambda$-distributional, where $\sqcup_\lambda$ is the Sugeno conorm defined by $x \sqcup_\lambda y := (x+y+\lambda xy) \wedge 1, x, y \in [0,1]$. If $\lambda = 0$ then $\nu_\lambda = \text{Pr}$. If $\lambda \neq 0$, then letting $K := \lfloor \log_2(n+1) \rfloor$, then $\mathcal{S}$ is g_λ-aggregable for some structural aggregation function g_λ iff

$$N \in I := \{1, 3, 7, \ldots, 2^{K-1} - 1, 2^K - 1\}.$$

In general, no numerical aggregators exist. Further, non-probabilistic Sugeno measures produce complete random sets only if $\exists k \in \{1, 2, \ldots\}, N = n = 2^k - 1$. Thus this broad class of distributional fuzzy measures does not generally yield aggregable random sets.

- Now consider an important class of aggregable random sets, and ask whether distributional fuzzy measures are forthcoming. In particular, consider ring-structured random sets with aggregation functions $g(A_j) := A_j \cap A_{j-1} = \{\omega_j\}$ where $A_0 = A_N$ by convention. If $\mathcal{S}$ is complete, then $h(m_j) = m_j + m_{j-1}$. But $\mathcal{S}$ is not distributional for any distribution operator $\sqcup$.

Thus in general probability and possibility stand as special cases which provide both distributional evidence measures and aggregable random sets. These results are summarized in Tables 4 and 4, and diagrammed in Fig. 3.

5 Possibilistic Processes

So far, we have motivated possibilistic systems theory within an overall GIT, also including imprecise probabilities, by first semantically grounding them in random interval measurement, and them justifying them as special distributional and aggregable forms. We now point the way to the other crucial aspect of a full systems and modeling theory necessary to complement measurement procedures, namely prediction methods. In particular, we introduce possibilistic processes as correlates to first-order Markov processes (see [13, 16], and [11]).

We can define a system which acts as a generalized first-order Markov process as a system $\mathcal{Z} := \langle S, \phi^0, V, \mathcal{R}, \Delta \rangle$ where S is a set of states; V is the valuation set, a lattice with $0, 1 \in V$ (here we assume

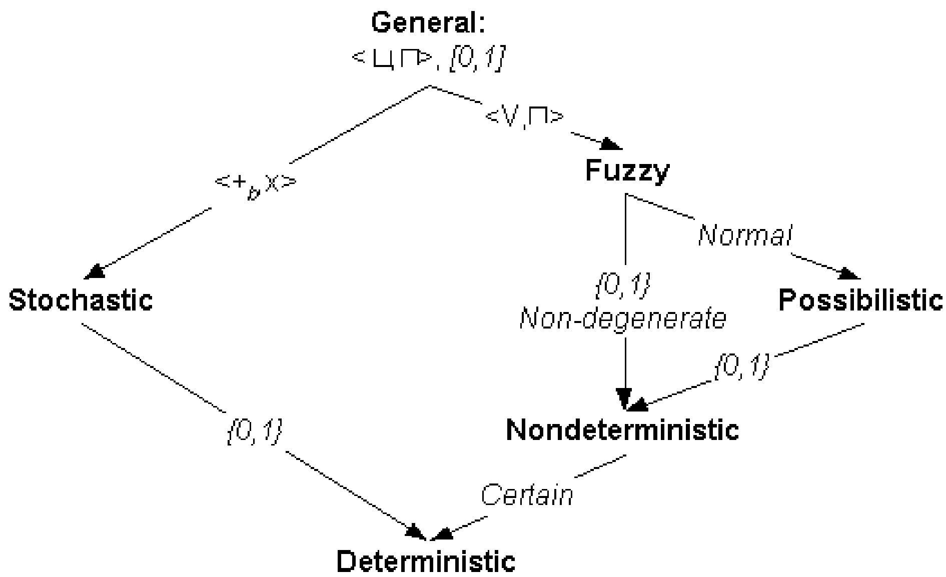

Figure 3: Relations among classes of processes.

	General	**Probabilistic**	**Possibilistic**
Topology	$2^\Omega - \{\emptyset\}$	Specific	Consonant
Distribution	$q_j = \mathrm{Pl}_j$	$p_j = \Pr(\{\omega_j\})$	$\pi_j = \Pi(\{\omega_j\})$
t-conorm	$\sqcup$	$+_b$	$\vee$
Focal Element	$A_j = g^{-1}(\omega_j)$	$\{\omega_j\}$	$\{\omega_1, \omega_2, \ldots, \omega_j\}$
Structural Aggregation	$g(A_j) = \omega_j$	A_j	$A_j - A_{j-1}$
Numerical Aggregation	$h(m_j) = \mathrm{Pl}_j$	m_j	$\sum_{k=j}^n m_j$
Inverse	$h^{-1}(\mathrm{Pl}_j) = m_j$	p_j	$\pi_j - \pi_{j+1}$
Completion	$\vert\mathcal{F}(\mathcal{S})\vert = \vert\Omega\vert$	$p_j > 0$	$\pi_j > \pi_{j+1}$

Table 1: Summary of the existing information theories.

	Sugeno	**Ring**
Topology	Sub-hypercube	Ring
t-conorm	$\sqcup_\lambda$	None
Focal Element	Any $\emptyset \neq A \subseteq \Omega$	$\{\omega_j, \omega_{j+1}\}$
Structural Aggregation	Only for $N \leq 2^{\lfloor \log_2(n+1) \rfloor} - 1$	$A_j \cap A_{j+1}$
Numerical Aggregation	None	$m_j + m_{j-1}$
Inverse	Never	For n even
Completion	$n = 2^k - 1$	$\mathrm{Pl}_j > 0$

Table 2: Summary of the special random set cases.

that V is a chain with $V \subseteq [0,1]$); $\mathcal{R} = \langle \sqcup, \sqcap \rangle$ is a conorm semiring; $\Delta: S^2 \mapsto V$ is the transition function; and $\phi^\tau: S \mapsto V$ are a family of state functions for $\tau \in \{0, 1, \ldots\}$, with ϕ^0 a given initial state function; and $\forall s \in S, \tau > 0$,

$$\phi^\tau(s) := \bigsqcup_{s' \in S} \phi^{\tau-1}(s') \sqcap \Delta(s, s'). \qquad (3)$$

When S is finite with $S = \{s_i\}, 1 \leq i \leq n := |S|$, then it is common to consider ϕ^τ as the vector $\vec{\phi}^\tau = \langle \phi_i^\tau \rangle$, with $\phi_i^\tau := \phi^\tau(s_i)$; Δ as a matrix $\Delta = [\Delta_{ij}]$ for $1 \leq i, j \leq n$, with $\Delta_{ij} := \Delta(s_i, s_j)$; and $\vec{\phi}^\tau = \vec{\phi}^{\tau-1} \circ \Delta$ where $\circ$ is matrix composition over the semiring $\mathcal{R}$, as shown in (3). Furthermore, ϕ^t is normal if $\bigsqcup_{s \in S} \phi^t(s) = 1$; Δ is transition normal if $\forall s' \in S, \bigsqcup_{s \in S} \Delta(s, s') = 1$; and $\mathcal{Z}$ is normal if Δ is transition normal and $\forall t \geq 0, \phi^t$ is normal. By theorem, if ϕ^0 is normal and Δ is transition normal, then $\mathcal{Z}$ is normal.

A number of cases follow depending on the specializations made for $\mathcal{R}, V$, and normalization, which are summarized in Tab. 3:

Stochastic Processes: Result when $\mathcal{R} = \langle +_b, \times \rangle$ is an additive semiring, so that the Δ_{ij} are the conditional probabilities of transiting from state s_j to state s_i and $\circ$ is normal matrix composition $\cdot$. Here normalization by $+$ is required, so that $\forall \tau, \sum_i p_i^\tau = 1$. This implies the weaker conorm $+_b$ normalization $(\sum_i p_i^\tau) \wedge 1 = 1$.

General Fuzzy Processes: Result when $\mathcal{R} = \langle \vee, \sqcap \rangle$ for any norm $\sqcap$. $\Delta \,\tilde{\subseteq}\, S^2$ is now a fuzzy matrix representing a fuzzy relation of the fuzzy linkage between the prior state s' and the subsequent state s; and $\circ$ is fuzzy matrix composition [26]. Note that there is no normalization, and all values are unconstrained over $[0, 1]$.

Nondeterministic Processes: If now V is restricted to $\{0, 1\} \subseteq [0, 1]$, then a classical nondeterminstic process results [10], so that at time τ there exists a set of possible states and any state can transit to multiple states.

Deterministic: Given either a stochastic process with $V = \{0, 1\}$, or a nondeterministic process with the certainty requirement $\forall \tau, \exists! s_i, \phi^t(s_i) = 1$, then a classical deterministic process results [10], which is always in one definite state, and transits to another definite state.

Possibilistic: Finally, given a fuzzy process which is normal by $\vee$, then a possibilistic process results [12]. Now $\pi^\tau(s_i) := \phi^\tau(s_i) \in [0, 1]$ is the possibility of being in state s_i at time τ; Δ is called a possibilistic matrix $\mathbf{\Pi} := \Delta$, with $\pi^\tau(s_i | s_j) := \mathbf{\Pi}_{ij} = \Delta_{ij}$ being the conditional possibility of transiting from state s_j to state s_i; and $\circ$ is fuzzy matrix composition.

6 Conclusion

We surveyed aspects of possibilistic systems theory in the context of GIT and imprecise probabilities. As the community moves to the articulation of a complete GIT involving these components and others, it will be important to consider them in mutual interaction, for the various strengths and benefits that each particular theory, or a general theory, can bring to particular interpretations and applications.

7 Acknowledgements

I would like to acknowledge both the critical and complementary comments of Peter Walley and the reviewers.

References

[1] de Cooman, Gert and Aeyels, Dirk: (1999) "A Random Set Description of a Possibility Measure and Its Natural Extension", *IEEE Trans. on Systems, Man and Cybernetics*, in press

[2] de Cooman, Gert; Ruan, D; and Kerre, EE, eds.: (1995) *Foundations and Applications of Possibility Theory*, World Scientific, Singapore

[3] de Cooman, Gert and Walley, Peter: (1999) *Imprecise Probabilities Project*, `http://ensmain.rug.ac.be/~ipp`

[4] Dempster, AP: (1967) "Upper and Lower Probabilities Induced by a Multivalued Mapping", *Annals of Mathematical Statistics*, v. **38**, pp. 325-339

[5] Dubois, Didier and Prade, Henri: (1982) "Class of Fuzzy Measures Based on Triangular Norms", *Int. J. of General Systems*, v. **8**, pp. 43-61

[6] Dubois, Didier and Prade, Henri: (1993) "Fuzzy Sets and Probability: Misunderstandings, Bridges and Gaps", in: *Proc. FUZZ-IEEE '93*

[7] Goodman, IR: (1994) "A New Characterization of Fuzzy Logic Operators Producing Homomorphic-Like Relations with One-Point Coverages of Random Sets", in: *Advances in Fuzzy Theory and Technology*, ed. PP Wang, pp. 133-159, Bookwrights Press, Raleigh NC

Class	Denotation	$\mathcal{R}$	V	Normal
Stochastic	$\mathcal{Z}_p$	$\langle +_b, \times \rangle$	$[0,1]$	By +
Fuzzy	$\tilde{\mathcal{Z}}$	$\langle \vee, \sqcap \rangle$	$[0,1]$	Not necessarily
Nondeterministic	$\mathcal{Z}_n$	$\langle \vee, \sqcap \rangle$	$\{0,1\}$	Yes
Deterministic	$\mathcal{Z}_d$	$\langle +_b, \times \rangle = \langle \vee, \sqcap \rangle$	$\{0,1\}$	Yes
Possibilistic	$\mathcal{Z}_\pi$	$\langle \vee, \sqcap \rangle$	$[0,1]$	Yes

Table 3: Special cases of processes.

[8] Goutsias, J; Mahler, Ronald PS; and Nguyen, HT, eds.: (1998) *Random Sets: Theory and Applications*, Springer-Verlag

[9] Guan, J and Bell, DA: (1992) *Evidence Theory and its Applications*, v. **1-2**, North Holland

[10] Hopcroft, John E and Ullman, Jeffery D: (1979) *Introduction to Automata Theory Languages and Computation*, Addison-Wesley, Reading MA

[11] Janssen, Hugo J; Cooman, Gert de; and Kerre, Etienne: (1996) "First Results for a Mathematical Theory of Possibilistic Processes", in: *Proc. 1996 European Meeting on Cybernetics and Systems Research*, pp. 341-346

[12] Joslyn, Cliff: (1994) "On Possibilistic Automata", in: *Computer Aided Systems Theory—EUROCAST '93*, ed. F. Pichler and R. Moreno-Diáz, pp. 231-242, Springer-Verlag, Berlin

[13] Joslyn, Cliff: (1994) "Possibilistic Approach to Qualitative Model-Based Diagnosis", *Telematics and Informatics*, v. **11**:4, pp. 365-384

[14] 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

[15] 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

[16] Joslyn, Cliff: (1996) "The Process Theoretical Approach to Qualitative DEVS", in: *Proc. 7th Conf. on AI, Simulation, and Planning in High Autonomy Systems (AIS '96)*, ed. Bernie Zeigler, pp. 235-242, San Diego CA

[17] Joslyn, Cliff: (1997) "Measurement of Possibilistic Histograms from Interval Data", *Int. J. General Systems*, v. **26**:1-2, pp. 9-33

[18] Joslyn, Cliff: (1997) "Possibilistic Normalization of Inconsistent Random Intervals", *Advances in Systems Science and Applications*, v. **special**, ed. Wansheng Tang, pp. 44-51, San Marcos, TX

[19] 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

[20] Joslyn, Cliff and Klir, George: (1992) "Minimal Information Loss Possibilistic Approximations of Random Sets", in: *Proc. 1992 FUZZ-IEEE Conference*, ed. Jim Bezdek, pp. 1081-1088, IEEE, San Diego

[21] Joslyn, Cliff and Rocha, Luis: (1998) "Towards a Formal Taxonomy of Hybrid Uncertainty Representations", *Information Sciences*, v. **110**:3-4, pp. 255-277

[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] Klir, George: (1984) "Possibilistic Information Theory", in: *Cybernetics and Systems Research*, v. **2**, ed. R. Trappl

[24] Klir, George: (1991) "Generalized Information Theory", *Fuzzy Sets and Systems*, v. **40**, pp. 127-142

[25] Klir, George: (1993) "Developments in Uncertainty Based Information", in: *Advances in Computers*, v. **36**, ed. M. Yovitz, pp. 255-332, Academic Press

[26] Klir, George and Yuan, Bo: (1995) *Fuzzy Sets and Fuzzy Logic*, Prentice-Hall, New York

[27] Meystel, Alex: (1996) "Intelligent Systems: A Semiotic Perspective", *Int. J. Intelligent Control and Systems*, v. **1**, pp. 31-57

[28] Pawlak, Zdzislaw: (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Boston

[29] Shafer, Glen: (1976) *Mathematical Theory of Evidence*, Princeton U Press, Princeton

[30] Walley, P: (1990) *Statstical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York

[31] Walley, P: (1996) "Measures of Uncertainty in Expert Systems", *Artificial Intelligence*, v. **83**, pp. 1-58

[32] Walley, P: (1996) "Inferences from Multinomial Data: Learning About a Bag of Marbles", *J. Royal Statistical Society B*, v. **58**, pp. 3-57

[33] Wang, Zhenyuan and Klir, George J: (1992) *Fuzzy Measure Theory*, Plenum Press, New York