Int. J. General Systems, Vol. 26(1–2), pp. 09–33
Reprints available directly from the publisher
Photocopying permitted by license only

MEASUREMENT OF POSSIBILISTIC HISTOGRAMS FROM INTERVAL DATA

CLIFF JOSLYN*

Los Alamos National Laboratory, Los Alamos, New Mexico 87543, U.S.A.

(Received January 26, 1996)

Measurement methods are a central requirement for the semantic grounding of any mathematical systems theory. Therefore possibility theory, as a branch of General Information Theory (GIT), requires objective measurement methods to extend its agenda and applications beyond the fuzzy theory from which it emerged. General measuring devices, when defined on intervals of $\mathbb{R}$, yield empirical random intervals which, when consistent, yield possibility distributions as their plausibilistic traces. These empirical possibility distributions are called possibilistic histograms, and are fuzzy intervals. Their continuous approximations, even for very small sample sizes, yield the the standard fuzzy interval forms commonly used in fuzzy system applications.

INDEX TERMS: Possibilistic histograms, semiotics, possibilistic measurement, general information theory, fuzzy measures, possibility theory, random sets, random intervals, fuzzy intervals, fuzzy numbers

INTRODUCTION: MEASUREMENT IN POSSIBILISTIC SEMIOTICS

Historically, science has moved from deterministic models to those, like quantum and statistical mechanics and chaotic dynamics, where uncertainty and indeterminism are represented in terms of probability theory. Stochastic descriptions have been linked to a rich theory of statistical information based on the stochastic entropy function. Systems Science has traditionally existed within this formal context, where distinctions are drawn between deterministic and nondeterministic systems, with nondeterminism represented by probability theory.

But recent years have seen a proliferation of new, non-probabilistic mathematical methods for the representation of uncertainty and information in systems models. Following Klir [1991], we call these methods collectively "General Information Theory" (GIT), which includes fuzzy sets, systems, and logic [Klir and Yuan, 1995]; fuzzy measures [Wang and Klir, 1992]; random set [Kendall, 1974] and Dempster-Shafer evidence theory [Dempster, 1967; Shafer, 1976]; possibility theory [de Cooman *et al.*, 1995]; imprecise probabilities [Walley, 1990]; probability bounds [Ferson *et al.*, 1996]; rough set theory [Pawlak, 1991]; and others. Each of these involves some form of generalization or extension away from stochastic representations.

Possibility theory is a component of GIT which has particular significance for Systems Science. This is because although probability and possibility are logically independent, they exist in parallel and are related within GIT in a formally analogous manner [de Cooman, 1995]: probability and possibility measures arise in Dempster-Shafer evidence theory as fuzzy

*Email: joslyn@lanl.gov, http://gwis2.circ.gwu.edu/~joslyn, (505) 667–9096.

measures defined on random sets; possibility and their dual necessity measures represent extreme ranges of probability intervals; and the distributions of all these generally non-additive fuzzy measures are fuzzy sets.

Possibilistic systems in particular provide an important class of generalizations of classical and stochastic systems. For example, possibility distributions generalize real intervals, possibilistic mathematics generalizes interval analysis, and possibilistic processes generalize non-deterministic processes. Possibilistic versions of the key components of stochastic systems theory, including automata, Markov processes, networks, and Monte Carlo methods, are all available [Janssen *et al.*, 1996; Joslyn, 1994a, 1994b, 1994c].

The Semiotic Requirement for Measurement in Systems Science

All of these components of GIT, including possibility theory, are in and of themselves mere mathematical formalisms, mere syntax. The usefulness of such theories, their semantics and meaning, only become evident when used within the broader context of general science, Systems Science, or engineering applications. It is all too easy to commit referential fallacies, mistaking the map for the territory by focusing completely on our symbol strings and losing sight of the underlying processes of measurement and interpretation.

We base the semantics and interpretation of possibility theory on ideas from semiotics about sign-functions and codings, and the isomorphic ideas from Systems Science about models. The field of semiotics [Deely, 1990; Eco, 1986] proposes three necessary elements for any symbol system: syntax, semantics and pragmatics. Syntactic relations are the formal, in our case mathematical, rules governing manipulation of the symbols themselves. Semantic relations link symbols to their meanings and interpretations. And finally pragmatic relations concern the use of the symbols in the world.

These ideas are expressed in Systems Science in terms of the modeling relation, shown in Figure 1 [Joslyn, 1995c; Rosen, 1985]. Assume sets $W = \{w\}$ and $M = \{m\}$; an object system $S_1 = \langle r, W\rangle$, $r\colon W \mapsto W$; a modeling system $S_2 = \langle f, M\rangle$, $f\colon M \mapsto M$; and a coding or measurement function $o\colon W \mapsto M$. Then $O := \langle S_1, S_2, o\rangle$ is a model if r, f, and o form a homomorphism, so that $\forall w \in W$, $o(r(w)) = f(o(w))$.

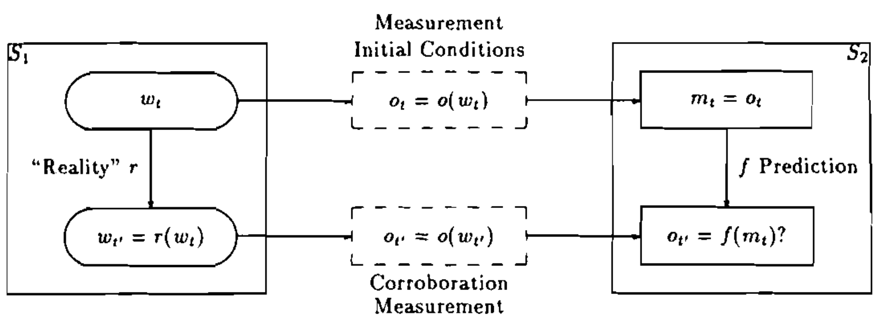

Figure 1 Models in natural science.

Semiotically, o is called a sign-function between the $w \in W$ and the $m \in M$, so that given $m = o(w)$, then m represents the referent w in virtue of the coding o. Thus the coding o serves a semantic function within the model, encoding the meaning of the w in terms of the m, while the functions r and f serve syntactic functions, transferring the w and the m through some dynamical processes into the future.

Natural science is concerned with the case where $S_1 = \langle W, r \rangle$ is a system of causal, ontological entailments, an aspect of the natural world. In this context the semantic coding function of o is understood as the actual physical *process* of measurement. The measurement $o_t = o(w_t)$ at time t is used to instantiate a model. Then the output of the model $m_{t'} = f(m_t) = f(o_t)$ at a different time t' is corroborated against the measurement $o_{t'} = o(w_{t'})$. If these "match", according to the appropriate criteria based on the nature of S_1 and S_2, then O is a good model.

Figure 1 shows a general modeling relation. But the formal properties of the modeling system S_2, the modeling language, necessarily constrain the set of possible interpretations of the object system S_1, and the figure must be modified or instantiated in particular ways as appropriate given the nature of S_2.

For example, if differential equations are used in S_2 as a modeling language, then the measurement methods used must produce states of the model m_t which are represented mathematically as states of a dynamical system, and the prediction function f must produce other such states, for example by integration. Similarly, if probability theory is used in S_2, then measurement methods must produce model states which are probability distributions, and prediction functions must be those from stochastic systems theory, for example a Markov process.

But still the relation between a given mathematical formalism and the measurement methods used to ground its symbols is necessarily nondeterminative and extratheoretical. Although the properties of S_2 *constrain* the possible interpretations, they do not *determine* them: each formalism might be left with multiple possible consistent interpretations. And certainly at the pragmatic level formalisms may be more or less appropriate for certain applications. Together, the pairing of a formalism with a particular interpretation and field of pragmatic applications delineates the "agenda" (to paraphrase Lotfi Zadeh[1]) of that field.

Existing Semiotics for Possibility Theory

Fuzzy systems theory is by far the most prevalent component of GIT, and until recently possibility theory has been tightly linked to fuzzy systems, both in its mathematics and its semantics [Joslyn, 1995a]. Thus it is natural that the fuzzy agenda has come to dominate not only possibility theory, but many of the other components of GIT as well.

Traditional fuzzy semantics is based on the interpretation of fuzzy sets as representations of human, cognitive categories. Measurement in fuzzy systems is traditionally based on cognitive modeling, usually of "linguistic variables". Similarly, its applications are traditionally in the engineering of human-created technological systems such as knowledge-based control systems, artificial intelligence and approximate reasoning systems, and decision support systems.

---

[1]Presentation at the First International Workshop on the Foundations and Applications of Possibility Theory, University of Ghent, 1995.

But there is nothing *a priori* necessary about this semiotics: GIT has other components, there are measurement methods other than cognitive modeling which are possible for these components, and there are applications in other than the typical "information engineering" domains. My purpose is to expand the semiotics of possibility theory beyond the traditional fuzzy semiotics to include the modeling of complex systems without regard to the presence of a human, cognitive agent. It is thus necessary to develop possibilistic measurement procedures based on objective, *empirical* observation.

There are, of course, existing empirical methods to derive measured possibility distributions in a more or less objective manner. These fall into three broad categories:

Knowledge Elicitation Measurement methods utilize the human subject in two ways. In a personalist approach, more or less complicated or sophisticated methods are used to determine the opinions of individuals about the membership grades of some state of affairs. Alternatively, a variety of methods are used to determine the values of membership grades of some cognitive state of a human subject. The former are subjective methods, and the latter may be objective, but both are too limiting for the possibilistic semiotics we wish to advance.

Converted Probabilities A variety of methods exist to convert a probability distribution, usually from an objectively measured frequency distribution, into a possibility distribution. While of course such transformations must be used when only frequency data are available, it can be shown that the resulting possibilistic representation is never ultimately appropriate for data initially governed by a specific frequency distribution [Joslyn, 1995b]. Specific data have very strong informational structures, much stronger than the very weak possibilistic structures. The difference between stochastic and possibilistic information is extraordinary: probability distributions provide virtually no possibilistic information, and thus virtually all conversions from frequency distributions yield incompatible possibility distributions.

Possibilistic Clustering Finally, we can identify fuzzy and possibilistic clustering methods [Barone and Filev, 1995; Krishnapuram and Keller, 1993] which construct possibility distributions based on a set of measured specific vectors. We find these methods to be an important contribution to the objective measurement of possibility distributions, but here propose an alternative strategy based on a collection of measured, non-specific, interval-valued observations.

A New Semiotics for Possibility Theory

When possibilistic data are desired, it is almost always preferable to obtain them in a form similar to their possibilistic representation. Thus objective, empirical measurement procedures are required that yield data in accordance with the semantic aspects of possibility theory, and thus governed by the mathematics of possibility theory.

The additivity of frequency data results from the specificity of observations of singletons, or indeed elements of any disjoint class. Therefore, the first step towards possibilistic measurement is the allowance for the possibility of *non-specific* data which are possibly *non-disjoint*, and thus *not* yielding traditional frequency distributions. This is essentially the concept of **set statistics**, originally advanced by Wang and Liu [Wang and Liu, 1984], and developed more by Dubois and Prade [Dubois and Prade, 1989, 1992].

In this paper we use interval valued set statistics collected from general measuring devices to develop a method for the objective measurement of possibility distributions in the form of possibilistic histograms. We first introduce the basics of mathematical possibility theory in terms of fuzzy measures, random set and evidence theory, fuzzy sets, and particularly random and fuzzy intervals. We then generalize classical, point-valued measurement in terms of general measuring devices on subsets, and consider special cases relevant for possibility theory. Measuring devices defined on intervals of $\mathbb{R}$ yield empirical random intervals, and when consistent, their plausibilistic traces are empirical possibility distributions called possibilistic histograms. We prove these to be fuzzy intervals. Finally, we show that continuous approximations to possibilistic histograms, even for very small sample sizes, yield the the standard fuzzy interval forms commonly used in fuzzy system applications.

Some aspects of this work have appeared in an introductory, synoptic, or unpublished form elsewhere [Joslyn, 1992, 1993b, 1994a].

MATHEMATICAL POSSIBILITY THEORY WITHIN GENERAL INFORMATION THEORY

We begin by introducing the basic aspects of mathematical possibility theory. This is done in the context of GIT, and so some related concepts of fuzzy sets, fuzzy measures, and random sets and intervals are also introduced.

Notation

Throughout the paper assume a universe of discourse $\Omega = \{\omega\}$. Generally consider $\Omega = \{\omega_i\}$, $1 \le i \le n$ to be finite. Sometimes we will explicitly recognize that $\Omega = \mathbb{R}$, and consider half-open interval subsets, elements of the class denoted $\mathcal{D} := \{[a, b) \subseteq \mathbb{R}: a, b \in \mathbb{R}, a < b\}$.

A vector denoted $\vec{a} = \langle a_i \rangle = \langle a_1, a_2, \ldots, a_m \rangle$ is a structure of length $|\vec{a}| := m$ where each element a_i of the vector is an element of some set $a_i \in X$. The a_i are ordered and may include duplicates. Let an element $b \in X$ be said to be included in a vector $b \in \vec{a}$ if $\exists a_i$, $b = a_i$. Define subtraction of an element a_i from a vector $\vec{a}$ as a new vector

$$\vec{a} - a_i := \langle a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_m \rangle$$

so that $|\vec{a} - a_i| = m - 1$.

Since a vector may contain duplicate elements $a_{i_1}, a_{i_2} \in \vec{a}$, $a_{i_1} = a_{i_2}$, therefore each vector $\vec{a}$ determines a unique non-empty set A constructed by including one instance of each element $a_i \in \vec{a}$, so that $b \in \vec{a} \leftrightarrow b \in A$, $1 \le |A| \le m$, and the quantity $|\vec{a}| - |A|$ is the number of elements of $\vec{a}$ which are duplicates.

Given a class $C = \{A\} \subseteq 2^{\Omega}$, define the core and support respectively as

$$\mathbf{C}(C) := \bigcap_{A \in C} A, \quad \mathbf{U}(C) := \bigcup_{A \in C} A.$$

Finally, let $\vee$ be the maximum and $\wedge$ the minimum operator.

Possibility and Other Fuzzy Measures

The most general mathematical basis for possibility theory currently is fuzzy measure and integral theory, as exemplified by Wang and Klir [1992] and de Cooman [1997a, 1997b, 1997c]. These researchers generalize the standard possibility theory by extending possibility measures to the positive reals and to lattices respectively. We will not consider these extensions directly here, except to briefly introduce some concepts of fuzzy measures.

The function $\nu\colon 2^{\Omega} \mapsto [0, 1]$ is a (finite) fuzzy measure [Wang and Klir, 1992] if $\nu(\emptyset) = 0$ and $\forall A, B \subseteq \Omega, A \subseteq B \rightarrow \nu(A) \leq \nu(B)$. Then $q_\nu\colon \Omega \mapsto [0, 1]$ with $q_\nu(\omega_i) := \nu(\{\omega_i\})$ is a distribution of ν if there exists a distribution operator function $\oplus\colon [0, 1]^2 \mapsto [0, 1]$ where $\langle [0, 1], \oplus, 0\rangle$ is an Abelian monoid ($\oplus$ is a commutative, associative, operator with identity 0) and, in operator notation,

$$\forall A \subseteq \Omega, \quad \bigoplus_{\omega_i \in A} q_\nu(\omega_i) = \nu(A).$$

Furthermore, ν is normal when $\nu(\Omega) = 1$, so that $\oplus_{\omega_i \in \Omega} q_\nu(\omega_i) = 1$. For a fixed finite fuzzy measure ν, denote $\vec{q} = \langle q_i \rangle := \langle q_\nu(\{\omega_i\})\rangle$ for $1 \leq i \leq n$.

Probability theory results from considering the fuzzy measure Pr with probability distribution $p := q_{\mathrm{Pr}}, \vec{p} = \langle p_i \rangle := \vec{q}$ where $\oplus = +$. The standard forms of probability result:

$$\forall A, B \subseteq \Omega, \quad \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B),$$

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

The central tenet of possibility theory is the introduction of a fuzzy measure Π with possibility distribution $\pi := q_\Pi$, $\vec{\pi} = \langle \pi_i \rangle := \vec{q}$. The equations of probability now take the form:

$$\forall A, B \subseteq \Omega, \quad \Pi(A \cup B) = \Pi(A) \vee \Pi(B),$$

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i, \tag{2}$$

$$\bigvee_{i=1}^{n} \pi_i = 1. \tag{3}$$

Random Set and Evidence Theory

One of the richest domains in GIT is that of random sets.

DEFINITION 4. (GENERAL RANDOM SET) Given a probability space $\langle \Omega, \Sigma, \Pr \rangle$, then a function $S\colon \Omega \mapsto 2^{\Omega} - \{\emptyset\}$, where $-$ is set subtraction, is a random subset of Ω if S is Pr-measurable, so that $\forall \emptyset \neq A \subseteq \Omega, S^{-1}(\mathbf{A}) \in \Sigma$.

Random sets were originally developed as a branch of stochastic geometry, and their mathematics in general can be quite complex [Artstein and Vitale, 1975; Kendall, 1974]. But for our purposes, and especially 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 [Dempster, 1967; Shafer, 1976]. We now reintroduce random sets in this context.

DEFINITION 5. (EVIDENCE FUNCTION, BASIC ASSIGNMENT) A function $m: 2^{\Omega} \mapsto [0, 1]$ is an evidence function (basic assignment) when $m(\emptyset) = 0$ and $\Sigma_{A \subseteq \Omega} m(A) = 1$.

DEFINITION 6. (FINITE RANDOM SET) Given an evidence function m, then

$$S := \{\langle A_j, m_j \rangle : m_j > 0\}, \tag{7}$$

is a finite random set where $A_j \subseteq \Omega$, $m_j := m(A_j)$, and $1 \leq j \leq N := |S| \leq 2^n - 1$. Denote the focal set of S as the class $\mathcal{F}(S) := \{A_j : m_j > 0\} \subseteq 2^{\Omega}$, and denote $\mathbf{C}(S) := \mathbf{C}(\mathcal{F}(S))$, $\mathbf{U}(S) := \mathbf{U}(\mathcal{F}(S))$.

DEFINITION 8. (RANDOM SET CONSISTENCY) A random set S is consistent if $\mathbf{C}(S) \neq \emptyset$.

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

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

where $A \perp B$ denotes $A \cap B = \emptyset$. Pl and Bel are generally non-additive fuzzy measures [Wang and Klir, 1992] without distributions, and are dual, in that $\forall A \subseteq \Omega$, $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\bar{A})$. The plausibility assignment (otherwise known as the trace or one-point coverage) of S is $\vec{\rho}(S) = \langle \rho_i \rangle$, where $\rho_i := \mathrm{Pl}(\{\omega_i\}) = \Sigma_{A_j \ni \omega_i} m_j$.

When Pl has a distribution operator $\oplus$, then $\vec{q}(S) := \vec{\rho}(S)$ is called the distribution of S. In particular, when

$$\forall A_j \in \mathcal{F}(S), \qquad |A_j| = 1, \tag{9}$$

then S is called specific, $\Pr(A) := \mathrm{Pl}(A) = \mathrm{Bel}(A)$ becomes a probability measure, and $\vec{p}(S) := \vec{q}(S) = \vec{\rho}(S)$ is a probability distribution. Similarly, S is called consonant ($\mathcal{F}(S)$ 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.[2] $\vec{\pi} := \vec{q}(S) = \vec{\rho}(S)$ is then a possibility distribution.

Fuzzy Sets and Their Relation to Random Sets

In fuzzy set theory, the characteristic or indicator function $\chi_A: \Omega \mapsto \{0, 1\}$ of a subset $A \subseteq \Omega$, where $\forall \omega_i \in \Omega$

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

is generalized to the membership function $\mu_{\tilde{F}}: \Omega \mapsto [0, 1]$ of a fuzzy subset denoted $\tilde{F} \mathrel{\tilde{\subseteq}} \Omega$. The value of $\mu_{\tilde{F}}(\omega_i)$ indicates the degree or extent to which $\omega_i \in \Omega$. As with discrete distributions, denote $\vec{\mu} = \langle \mu_i \rangle := \langle \mu_{\tilde{F}}(\omega_i) \rangle$.

It is clear that a fuzzy measure distribution q_ν is the membership function of some fuzzy set, and in particular probability p and possibility distributions π are. Similarly, plausibility assignments $\vec{\rho}(S)$ of discrete random sets are the membership functions of fuzzy sets. Thus each random set S maps to a unique fuzzy set $\vec{\rho}(S)$, or to its distribution $\vec{q}(S)$ if $\oplus$ exists. But when we begin with a particular fuzzy set $\vec{\mu}$, there is generally a non-empty, nonunique equivalence class of random sets $\Psi(\vec{\mu})$ for which $\forall S \in \Psi(\vec{\mu})$, $\vec{\rho}(S) = \vec{\mu}$[Goodman, 1994].

---

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

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 (9)) random set. But when $\vec{\mu}$ begins as a maximal possibility distribution $\vec{\pi}$, then in general $|\Psi(\vec{\pi})| > 1$. All of the $S \in \Psi(\vec{\pi})$ are consistent, and thus it is the consistency condition of Definition (8) which is both necessary and sufficient for S to have a maximally normalized possibility distribution $\vec{\pi} = \vec{\rho}(S)$ by (3).

LEMMA 10. (NORMALIZATION AND CONSISTENCY) S is consistent iff $\bigvee_{i=1}^{n} \rho_i = 1$.

Proof **Case 1***:* Assume S is consistent. Then $\exists \omega_i \in \mathbf{C}(S)$, and therefore $\forall A_j \in \mathcal{F}(S)$, $\omega_i \in A_j$ and $\rho_i = \sum_{A_j \in \mathcal{F}(S)} m_j = 1$. **Case 2***:* Assume $\bigvee_i \rho_i = 1$ so that $\exists \omega_0 \in \Omega$, $Pl(\{\omega_0\}) = \sum_{\omega_0 \in A_j} m_j = 1 = \sum_j m_j$, so that it must be that $\forall A_j \in \mathcal{F}(S)$, $\omega_0 \in A_j$, *and thus* $\omega_0 \in \mathbf{C}(S)$, *so that S is consistent.* ∎

Consistency is weaker than consonance: each consonant random set is consistent with core $\mathbf{C}(S) = A_1$, but not vice versa. So while a consistent but non-consonant random set has a maximal plausibility assignment $\vec{\pi}(S) = \vec{\rho}(S)$, its plausibility measure Pl is *not* a possibility measure Π. But for a given possibility distribution $\vec{\pi}$, there is a unique consonant random set $S^*(\vec{\pi}) \in \Psi(\vec{\pi})$, and of course $\vec{\pi}$ is its possibility distribution. The possibility measure of $S^*(\vec{\pi})$, denoted Π^*, is constructed by invoking (2) on $\vec{\pi}$, and $S^*(\vec{\pi})$ can be constructed in turn from Π^* [Joslyn, 1993a, 1994a].

Π^* is the possibilistic approximation to the plausibility measures Pl of all the other $S \in \Psi(\pi)$, $S \neq S^*(\vec{\pi})$. Therefore, in general when working with possibility theory in the context of finite random sets, a consistent random set S is a sufficient condition to generate a possibility distribution $\vec{\pi}(S)$, with the knowledge that the unique approximating possibility measure Π^* and consonant random set $S^*(\vec{\pi}(S))$ are always available.

Random and Fuzzy Intervals

As we move to discuss possibilistic measurement proper, it will be desirable to let $\Omega = \mathbb{R}$.

DEFINITION 11. (RANDOM INTERVAL) 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 $\mathbb{R}$. For random intervals, we modify the concept of the plausibility assignment slightly.

DEFINITION 12. (RANDOM INTERVAL TRACE) Given a random interval $\mathcal{A}$, define the function $\rho_{\mathcal{A}}: \mathbb{R} \mapsto [0, 1]$ as the plausibilistic trace, or just trace, of $\mathcal{A}$, where

$$\forall x \in \mathbb{R}, \quad \rho_{\mathcal{A}}(x) := \mathrm{Pl}(\{x\}) = \sum_{A_j \ni x} m_j. \tag{13}$$

Note that $\rho_{\mathcal{A}}$ is the membership function of a fuzzy subset of $\mathbb{R}$. There are two special fuzzy subsets of $\Omega = \mathbb{R}$ which are of particular interest to us.

DEFINITION 14. (FUZZY INTERVAL) [Dubois and Prade, 1978, 1987] A fuzzy subset of the real line $\tilde{F} \tilde{\subseteq} \mathbb{R}$ is a **fuzzy interval** if $\tilde{F}$ is maximally normalized and convex, so that

$$\forall x, y \in \mathbb{R}, \quad \forall z \in [x, y], \quad \mu_{\tilde{F}}(z) \geq \mu_{\tilde{F}}(x) \wedge \mu_{\tilde{F}}(y).$$

DEFINITION 15. (FUZZY NUMBER) A **fuzzy number** is a fuzzy interval $\tilde{F}$ where $\exists x \in \mathbb{R}$, $\mathbf{C}(\tilde{F}) = \{x\}$.

NOTE 16. Fuzzy intervals generalize crisp intervals as fuzzy sets generalize crisp sets. Further, all fuzzy intervals and numbers are possibility distributions on $\Omega = \mathbb{R}$.

NOTE 17. Our definition differs somewhat from others in the literature [Gil, 1992] who use closed intervals. First, this choice makes some of the algebraic manipulations easier. But also note that $\mathcal{D}$ is the basis for constructing Borel σ-fields and Borel sets, and so our usage is more consistent with that of measure theory [Halmos, 1950; Wang and Klir, 1992].

NOTE 18. Previously Ω had been postulated as a finite set, which leads to a great deal of mathematical simplicity. However, even though Ω is now uncountable, complications can still be avoided as long as $\mathcal{A}$ is finite, that is as long as only finitely many (N) focal elements are present. This is because an interval $A = [a, b) \subseteq \mathbb{R}$ can be characterized completely by the two endpoints a and b. With each new focal element A_j, N grows by 1, and the total number of endpoints present in $\mathcal{F}(\mathcal{A})$ grows by at most 2. Thus the focal set of a finite random interval can be completely represented by the finite collection of these endpoints. It is only these endpoints that need to be considered, and none of the properties of the continuum of points between them is significant. This will be considered more completely in Theorems (39) and (41) below.

MEASURING DEVICES AND EMPIRICAL RANDOM SETS

Measurement is the general process of encoding an aspect of the "real world" into its representation in a formal system. It is only through measurement procedures that we can gain knowledge about the world; it is through the results of measurements that the world is "presented" to us.

Classical Measuring Devices

We generally think of a measuring device as producing a measured value which is a number $x \in \mathbb{R}$. For example, a thermometer calibrated in integral degrees in the interval [0, 100] would yield a result, say 72 degrees, $72 \in \{0, 1, \ldots, 100\}$. We call such a device a **classical measuring device**.

DEFINITION 19. (CLASSICAL MEASURING DEVICE) A **classical measuring device** is a system $\mathcal{M}_C := \langle \Omega', \vec{\omega}', c \rangle$, where:

- $\Omega' = \{\omega'_{i'}\}$, $1 \leq i' \leq n'$ is the set of **possible observations**;
- The vector $\vec{\omega}' := \langle \omega'_s \rangle$ is the **classical measurement record**, a list of all the points which were actually observed, with $1 \leq s \leq M$; and
- $c{:}\Omega' \mapsto \mathcal{W}$ is the **count function**, where $\mathcal{W}$ are the whole numbers and $\forall \omega'_{i'} \in \Omega'$, $c_{i'} := c(\omega'_{i'})$ is the count of the number of times $\omega'_{i'}$ is observed in $\vec{\omega}'$.

DEFINITION 20. (FREQUENCY DISTRIBUTION) Given a classical measuring device $\mathcal{M}_C$, then a **frequency distribution** is a function $f:\Omega' \mapsto [0, 1]$ where

$$f(\omega'_{i'}) = f_{i'} := \frac{c_{i'}}{\sum_{i'=1}^{n'} c_{i'}}.$$

Denote the vector $\vec{f} := \langle f_{i'} \rangle$.

DEFINITION 21. (FREQUENCY MEASURE) Given a frequency distribution f, then the **frequency measure** is a function $P:2^{\Omega'} \mapsto [0, 1]$ where $\forall A \subseteq \Omega'$, $P(A) := \sum_{\omega'_{i'} \in A} f_{i'}$.

$\vec{f}$ is a natural probability distribution, and P is a natural probability measure as in (1).

General Measuring Devices

On closer examination, however, we recognize that there is always some uncertainty on the readout of the thermometer. The thermometer is in fact a glass tube, a continuous object, which we represent as $\Omega \subseteq \mathbb{R}$. The tube is marked at certain points, say d_j, indicating the number of degrees. When the thermometer equilibrates, the mercury stops at some point almost always *between* two of the marked points.

While we can use subjective estimation to interpolate between these two points, within the formalism (or for a digital, electronic thermometer) only an *interval*, say $B_s := [d_s, d_{s+1})$, can be reported as the result of the measurement. While any particular interval B_s is usually identified by and reported as a single number, for example d_s, d_{s+1}, or the midpoint $\frac{d_s+d_{s+1}}{2}$, it must always be kept in mind that it in fact indicates the *entire* interval $[d_s, d_{s+1})$. Observation of a specific position of the mercury (an $x \in B_s$) must yield an interval readout B_s. Thus observation of only the interval B_s leaves uncertainty as to the "actual" value $x \in B_s$.

This leads to the concept of a **general measuring device**.

DEFINITION 22. (GENERAL MEASURING DEVICE) A **general measuring device** is a system $\mathcal{M} := \langle \mathcal{C}, \vec{B}, C \rangle$, 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 list 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 $\vec{B}$.

$\mathcal{C}$ is the collection of all subsets which are observable by the device, and is defined here as a finite class. Each time a measurement is taken, an $A_{j'} \subseteq \Omega$ results as a report of $\mathcal{M}$. The nature of the measuring device will depend on the elements and topological structure of $\mathcal{C}$. In the thermometer example, $\mathcal{C} = \{B_s\}$ is the collection of disjoint, equal length, half-open intervals $B_s = [d_s, d_{s+1})$.

Empirical Random Sets

Since $\vec{B}$ is a vector, it may contain duplicates; in other words, it may be that an element of $\mathcal{C}$ is observed more than once.

DEFINITION 23. (EMPIRICAL FOCAL SET) Given a general measuring device $\mathcal{M}$, let $\mathcal{F}^E := \{A_j\} \subseteq C$, $1 \leq j \leq N$, be an **empirical focal set** derived by eliminating the duplicates from $\vec{B}$, where:

$$1 \leq j \leq N, \quad \mathcal{F}^E \subseteq C, \quad N \leq N', \quad N \leq M, \quad \forall A_j \in \mathcal{F}^E, \exists B_S \in \vec{B}, A_j = B_s.$$

$\mathcal{F}^E$ is essentially the restriction of C to those subsets which are actually observed in the record $\vec{B}$. Each of the $A_{j'}$ is one of the N' observable sets; each of the B_s is one of the M records that an $A_{j'}$ has been observed; and finally each of the A_j is one of the N sets which was actually observed at one time or another. If $N < N'$, then not all observable sets are in fact observed; if $N < M$ then some observable sets were observed more than once. Denote C_j as that $C_{j'}$ for which $A_j = A_{j'}$, so that the C_j are the positive $C_{j'}$.

Now proceed in a manner analogous to frequency distributions in classical devices from Definition (20), but with set-valued observations.

DEFINITION 24. (SET-FREQUENCY DISTRIBUTION) Given a general measuring device $\mathcal{M}$ then a **set-frequency distribution** is a function m^E: $\mathcal{F}^E \mapsto [0, 1]$ where

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

COROLLARY 25. m^E is an evidence function.

Proof First, since $\sum_{j=1}^N C_j = M$, therefore

$$\sum_{j=1}^N m_j^E = \sum_{j=1}^N \frac{C_j}{M} = \frac{\sum_{j=1}^N C_j}{M} = M/M = 1.$$

Then, since $\emptyset \notin C$, therefore from Definition (5) m^E is an evidence function. ■

DEFINITION 26. (EMPIRICAL RANDOM SET) Given a general measuring device $\mathcal{M}$, let the **empirical random set** S^E be defined as in Definition (7) from $\mathcal{F}^E$ and m^E.

NOTE 27. Set-based statistics and empirically derived random sets have only a small presence in the literature. Within GIT they have been used primarily by Wang and Liu [1984, 1986] and Dubois and Prade [1989, 1990, 1992]. Fung and Chong [1986] provide an interesting example of the use of set statistics in their critique of Dempstere's rule of combination.

NOTE 28. There is another entirely different sense of "set-valued statistic", as used by Degroot and Eddy [1983]. Here it does not mean a mathematical property of some measured subset data, but rather an indeterminate value for the parameter of a probability distribution. For example, a uniform probability distribution would have a set-valued parameter if it was defined on a disconnected subset of the line, for example $[1, 2] \cup [5, 6]$.

Disjoint Measuring Devices

The key feature of a classical instrument is that its observable sets are *disjoint*.

DEFINITION 29. (DISJOINT MEASURING DEVICE) A general measuring device $\mathcal{M}$ is **disjoint** if $\forall A_1, A_2 \in C, A_1 \perp A_2$.

Generally, scientists strive to construct disjoint measuring devices. In such devices C is an equivalence class on Ω, establishing observations of $\omega \in \Omega$ in an equivalence relation. Furthermore, when C is a partition, that is $\cup_{j'=1}^{N'} A_{j'} = \Omega$, then C covers Ω, yielding all observations possible. Alternatively, even when C does not cover Ω, C does cover the sub-universe $(\cup_{j'=1}^{N'} A_{j'}) \subseteq \Omega$.

For a disjoint device, as with all general measuring devices, observations result in reports of $A_{j'} \in C$, not $\omega \in \Omega$. Thus the observation of an $A_{j'}$ leaves uncertainty as to a specific point outcome $\omega \in A_{j'}$. But what must be stressed is that *at the level of description of* C there is no uncertainty or ambiguity. Rather, the cardinalities $|A_{j'}|$ relative to $|\Omega|$ indicate the precision of the instrument.

So in this case C can itself be considered as a *new* universe of discourse $\Omega' := C = \{A_{j'}\}$. Of course Ω' is essentially equivalent to Ω. The difference is just that in Ω' the ω are grouped into the sets $A_{j'}$, and Ω' is considered as a collection of the $A_{j'}$, not of the ω.

Because the $A_{j'}$ are disjoint, so will the actual observed subsets A_j. Then $\vec{B}$ becomes a time-series data set on points in Ω', and the empirical evidence function m^E becomes a simple frequency distribution over the disjoint $A_{j'}$, as a true probability distribution, and not as an evidence function.

Thus we arrive at the following proposition.

PROPOSITION 30. If $\mathcal{M}$ is disjoint, then $\mathcal{M}$ is a classical measuring device $\mathcal{M}_C$ with $\Omega' = C$, $\vec{\omega}' = \vec{B}$, and $c = C$.

So we see that time-series data derived from measurement on classical instruments necessarily generate probability distributions. As argued above, a frequency conversion $f' \mapsto \pi$ can be constructed, but it is better to continue the search for appropriately possibilistic measured data.

Consistent and Consonant Measuring Devices

These are similar to consistent and consonant random sets.

DEFINITION 31. (CONSISTENT AND CONSONANT MEASURING DEVICES) A general measuring device $\mathcal{M}$ is consistent if C is consistent. A general measuring device $\mathcal{M}$ is consonant if C is a nest.

It it clear that consistent and consonant measuring devices yield consistent and consonant empirical random sets respectively. Recall that $\mathcal{F}^E = \{A_j\} \subseteq \{A_{j'}\} = C$.

COROLLARY 32. If $\mathcal{M}$ is consistent then S^E is consistent.

Proof Since $\mathcal{F}^E \subseteq C$ and $\forall A_{j'} \in C, \emptyset \neq \mathbf{C}(C) \subseteq A_{j'}$, therefore $\forall A_j \in \mathcal{F}^E$, $\mathbf{C}(C) \subseteq A_j$, so that $\mathbf{C}(C) \subseteq \mathbf{C}(\mathcal{F}^E)$, requiring $\mathbf{C}(\mathcal{F}^E) \neq \emptyset$. ■

COROLLARY 33. If $\mathcal{M}$ is consonant then S^E is consonant.

Proof Follows directly from $\mathcal{F}^E \subseteq C$. ■

Elsewhere I have described a number of situations which are best modeled as general measuring devices because of the presence of non-disjoint, possibly overlapping set-valued observations [Joslyn, 1992, 1994a, 1994b]. Here I would like to suggest a thought experiment for the physical realization of consistent and consonant measuring devices.

Consider Figure 2. There is a length of wall against which an experimenter can throw a ball, striking it somewhere between points a and b. On the left is a consistent device $\mathcal{M}$ for measuring the ball hits. Two observers O_1 and O_2 view the wall through two holes in a screen configured so that O_1 sees only the area $A_{1'}$, O_2 only the area $A_{2'}$, and there is a partial overlap. Similarly, on the right of the figure is a consonant device, where O_1 and O_2 look through the same hole in the screen, but O_2 is farther away. Now O_2 can see everything that O_1 sees, but not vice versa, so that $A_{1'} \subseteq A_{2'}$. In each case a number of balls are tossed, and each observer reports to the experimenter the total number of balls she saw hit.

Two points are crucial here. First, the experimenter has no independent knowledge as to the position where the balls hit except as reported by the observers. Thus while there may be a "real" position of the ball hits against the wall, all that is known is whether it hit in $A_{1'}$ or $A_{2'}$.

Second, the records of the hits reported by each observer cannot *correlated*, rather only a *statistical* description of the collection of observations, in this case of the total number of hits seen by each observer, can be reported. If the experimenter knew for any *particular* ball toss which of the observers (or both) reported, then he would be able to disambiguate where the ball struck within an element of a partition of the support $\mathbf{U}(\mathcal{C})$.

For example, for the consistent device if it was known for a particular ball throw that O_1 but not O_2 made a report, then the strike could be localized to the region $A_{1'} - A_{2'}$, identified as G_1 in the figure. The partition in this case is $\{G_1, G_2, G_3\}$. In the consonant case, if it was known for a particular ball throw that O_1 reported, then necessarily O_2 must also report. In this case the partition is $\{G_1 \cup G_3, G_2\}$. These cases are outlined in Table 1.

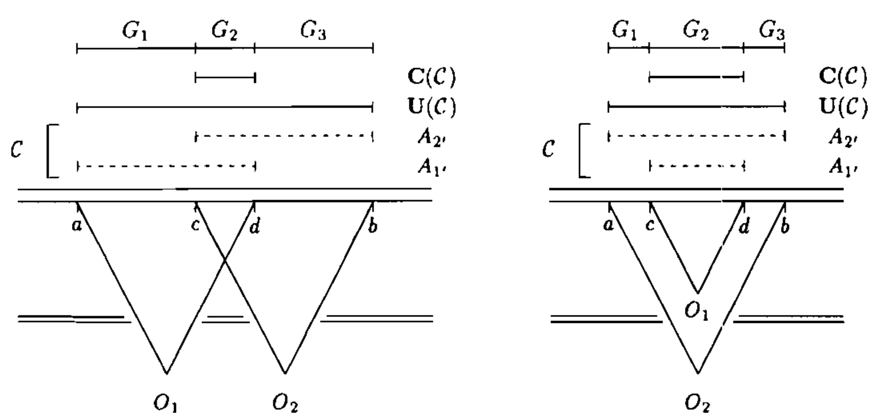

Figure 2 (Left) A consistent measuring device. (Right) A consonant measuring device.

Table 1 Results of correlated reports in the consistent and consonant example devices.

Device	Reports	Equivalence Class	Observable Set
Consistent	O_1, O_2	G_2	$A_{1'} \cap A_{2'}$
	$O_1, \neg O_2$	G_1	$A_{1'} - A_{2'}$
	$\neg O_1, O_2$	G_3	$A_{2'} - A_{1'}$
Consonant	O_1, O_2	G_2	$A_{1'}$
	$O_1, \neg O_2$	Impossible	Impossible
	$\neg O_1, O_2$	$G_1 \cup G_3$	$A_{2'} - A_{1'}$

POSSIBILISTIC HISTOGRAMS FROM EMPIRICAL RANDOM INTERVALS

We now move to possibilistic histograms by considering measuring devices which yield empirical random intervals.

DEFINITION 34. (INTERVAL MEASURING DEVICE) A measuring device $\mathcal{M}$ is an interval device $\mathcal{M}_I$ if $\Omega = \mathbb{R}$ and $C \subseteq \mathcal{D}$.

Note that now C may be infinite, either countably or uncountably, but as discussed in Note (18), we will restrict ourselves to finite samples of C.

DEFINITION 35. (EMPIRICAL RANDOM INTERVAL) For an interval measuring device $\mathcal{M}_I$ let the empirical random set S^E produced from the (finite) measurement record of $\mathcal{M}_I$ be an empirical random interval denoted $\mathcal{A}^E$.

In the sequel we will deal almost exclusively with random intervals both in general and in the consistent case.

PROPOSITION 36. From the plausibility assignment formula (13) and the set-frequency distribution function definition (24), given an empirical random interval $\mathcal{A}^E$, then $\forall x \in \mathbf{U}(\rho_{\mathcal{A}^E})$,

$$\rho_{\mathcal{A}^E}(x) = \sum_{A_j \ni x} m_j^E = \frac{\sum_{A_j \ni x} C_j}{M}.$$

The Form of Finite Random Intervals

The following definitions are summarized in Table 2, and are illustrated in the example in the next section using Figure 3.

First, it will prove very useful to denote the endpoints and the "order statistics" [David, 1981] of the endpoints of the focal elements of a random interval.

DEFINITION 37. (RANDOM INTERVAL FOCAL SET COMPONENTS) Assume a random interval $\mathcal{A}$.

- Denote by $A_j \in \mathcal{F}(\mathcal{A})$ the closed intervals $A_j := [l_j, r_j)$.
- Let $l_{(j)}$ and $r_{(j)}$ be the **order** and "**reverse order**" statistics of the left and right endpoints, so that

$$l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(N)}, \quad r_{(N)} \leq r_{(N-1)} \leq \cdots \leq r_{(1)}. \tag{38}$$

are permutations of the l_j, r_j.

Table 2 Summary of the components of the focal set and plausibilistic trace of a random interval.

Group	Components	Bound[a]	Description
$\mathcal{C}$	$\{A_j'\}$	N'	Observable class
$\vec{B}$	$\langle B_s \rangle$	M	Measurement record
$\mathcal{F}^E$	$\{A_j\}$	N	Empirical focal set
$\vec{E}^l, \vec{E}^r$	$\langle l_j \rangle, \langle r_j \rangle$	N	Left and right endpoint vectors
$\vec{E}$	$\langle l_j, r_j \rangle$	$2N$	Joint endpoint vector
$\hat{E}$	$\langle l_{(j)}, r_{(j)} \rangle$	$2N$	Ordered joint endpoint vector
E	$\{e_k\}$	Q	Endpoints set
E^l, E^r	$\{e^l_{k^l}\}, \{e^r_{k^r}\}$	Q^l, Q^r	Left and right endpoints
Γ	$\{G_k\}$	$Q-1$	Domain interval of ρ
Υ	$\{T_k\}$	$Q-1$	Function intervals of ρ

[a]Where $N \le N'$, $N \le M$, $N+1 \le Q \le 2N$, $1 \le Q^l, Q^r \le N$, and $Q \le Q^l + Q^r$.

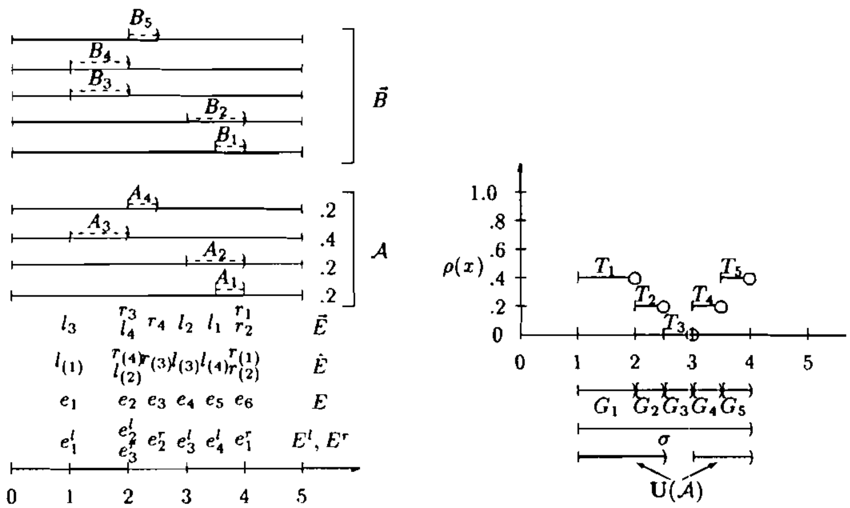

Figure 3 (Left) (top) An example measurement record from an interval measuring device on [0, 5). (bottom) Its random interval and order statistics. (Right) Its plausibilistic trace.

- Denote the vectors of **endpoints** and **ordered endpoints** as

$$\vec{E}^l := \langle l_1, l_2, \ldots, l_N \rangle, \quad \vec{E}^r := \langle r_1, r_2, \ldots, r_N \rangle, \vec{E} := \langle l_1, l_2, \ldots, l_N, r_1, r_2, \ldots, r_N \rangle.$$

$$\hat{E} := \langle l_{(1)}, l_{(2)}, \ldots, l_{(N)}, r_{(N)}, r_{(N-1)}, \ldots, r_{(1)} \rangle.$$

- Let

$$E^l := \{e^l_{k^l}\}, \quad E^r := \{e^r_{k^r}\}, \quad E := \{e_k\}.$$

be the sets of endpoints with duplicates omitted from $\vec{E}^l$, $\vec{E}^r$, and $\vec{E}$ respectively, where

$$\forall e^l_{k^l} \in E^l, e^l_{k^l} \in \vec{E}^l, \quad \forall e^r_{k^r} \in E^r, e^r_{k^r} \in \vec{E}^r, \quad \forall e_k \in E, e_k \in \vec{E},$$

$$1 \le k^l \le Q^l := |E^l|, \quad 1 \le k^r \le Q^r := |E^r|, \quad 1 \le k \le Q := |E|.$$

so that $E^l \cup E^r = E$ and $Q^l + Q^r \ge Q$. Further, let the $e_k \in E$ be ordered with $e_1 < e_2 < \cdots < e_Q$.

- Let $\Gamma := \{G_k\}$ for $1 \le k \le Q-1$, where $G_k := [e_k, e_{k+1})$.

THEOREM 39. For a random interval $\mathcal{A}$

$$1 \leq Q^r, Q^l \leq N, \quad N+1 \leq Q \leq 2N.$$

Proof The inequalities in (38) will be strict or not depending on whether a pair A_{j_1}, A_{j_2} share an endpoint. All the A_j are distinct, so they cannot share both endpoints. This forces most, but not all, of the l_j, r_j to be distinct. Consider an initial observation $A_1 := [a, b)$, so that now $N = 1$, $Q^l = Q^r = 1$, $Q = 2$. Then add a second observation $A_2 := [c, d)$, so that $N = 2$. One possibility is that $c \notin \{a, b\} \not\ni d$, so that now $Q^l = Q^r = 2$, $Q = 4$. On the other hand, if $c \in \{a, b\}$, then $d \notin \{a, b\}$, so $Q^l = 1$, $Q^r = 2$; and similarly if $d \in \{a, b\}$, then $c \notin \{a, b\}$, so $Q^l = 2$, $Q^r = 1$. In each case, $Q = 3$. So in general, as each observation is added, at least one, and at most two distinct endpoints are added. Thus the result follows. ■

We now characterize the plausibility assignments of random intervals. Below, denote $\rho := \rho_{\mathcal{A}}$ for a fixed random interval $\mathcal{A}$.

DEFINITION 40. (Plausibilistic Trace Form) Assume a random interval $\mathcal{A}$.

- For an interval $A \in \mathcal{D}$ and $y \in [0, 1]$, let $\rho(A) = y$ denote that $\forall x \in A$, $\rho(x) = y$.
- Let $\Upsilon := \{T_k\}$ for $1 \leq k \leq Q - 1$, where

$$T_k := \{\langle x, y \rangle \in G_k \times [0,1] : x \in G_k, y = \rho(x)\}.$$

  for $1 \leq k \leq Q - 1$.

ρ is a piecewise-constant collection of the left-closed segments T_k. It is thus completely characterized by the G_k, which partition the closure of the support $\mathbf{U}(\mathcal{A})$, and the values of ρ which are constant across each G_k. Further, as the ordinate x moves rightward, if it transits an endpoint which is only a left endpoint, ρ jumps discretely up; and as it transits an endpoint which is only a right endpoint, ρ jumps discretely down.

THEOREM 41. Assume a random interval $\mathcal{A}$.

1. If $\mathbf{U}(\mathcal{A})$ is not connected, then Γ contains a partition of $\mathbf{U}(\mathcal{A})$;
2. If $\mathbf{U}(\mathcal{A})$ is connected, then Γ is a partition $\mathbf{U}(\mathcal{A})$;
3. $\forall 1 \leq k \leq Q - 1, \quad \exists! y \in [0, 1], \quad \rho(G_k) = y, T_k = G_k \times \{y\}$;
4. $\forall 1 \leq k \leq Q - 2$:

$$e_{k+1} \in E^l, e_{k+1} \notin E^r \quad \to \quad \rho(G_k) < \rho(G_{k+1}),$$

$$e_{k+1} \in E^r, e_{k+1} \notin E^l \quad \to \quad \rho(G_k) > \rho(G_{k+1}).$$

Proof Recall that $\forall x \in \mathbb{R}$, $\rho(x) = \Sigma_{A_j \ni x} C_j / M$ and $\forall C_j > 0$. Thus $\rho(x)$ is proportional to the sum of the counts on all the focal elements A_j in which x is contained. Let $\sigma := [e_1, e_Q)$.

1. It is apparent that, for example, $G_1 \cup G_2 = [e_1, e_2) \cup [e_2, e_3) = [e_1, e_3)$, and so on, so that $\bigcup_{G_k \in \Gamma} G_k = \sigma$. Since the G_k are disjoint, therefore they partition σ. If $\mathbf{U}(\mathcal{A})$ is not connected, then there are some $x \in \sigma$ which for which $\rho(x) = 0$. These points exist in their

own (perhaps multiple) G_k, such that $\forall A_j \in \mathcal{F}(\mathcal{A}), A_j \perp G_k$. Remove all such G_k from Γ. The remaining G_k still partition the remaining portion of σ, and that portion is just $U(\mathcal{A})$.

2. Continuing from above, if $U(\mathcal{A})$ is connected, then no such G_k must be removed from Γ, and the G_k still partition $\sigma = U(\mathcal{A})$.
3. For any k, consider $x, x' \in G_k$. Since there is no endpoint e_{k_0} between x and x', therefore x and x' are contained within exactly the same collection of A_j, so that $\rho(x) = \rho(x') = y$ for some $y \in [0, 1]$. This is true $\forall x, x' \in G_k$, so that $y = \rho(G_k)$.
4. Select a G_k, select an $x \in G_k$, and select an $x' \in G_{k+1}$. Let $e_{k+1} = l_0 \in E^l$ be some left endpoint l_0, but not a right endpoint, so that $e_{k+1} \notin E^r$. Now l_0 is the only endpoint between x and x', so $\forall A_j \in \mathcal{F}(\mathcal{A}), l_j \leq x \rightarrow l_j \leq x'$, but $x < l_0 \leq x'$. So x' is contained within all of the A_j which contain x, but is also contained within $A_0 = [l_0, r_0)$. Therefore, from (13), $\rho(x') - \rho(x) = C_0/M > 0$, so that $\rho(G_k) = \rho(x) < \rho(x') = \rho(G_{k+1})$. A similar argument proves the final case.

■

Thus each of the e_k is equal to at least one of the (left or right) observed endpoints l_j or r_j, and ρ is completely determined by the coordinates $\langle e_k, \rho(e_k)\rangle$.

An Example

An example measurement record from an interval device, its random interval, and its plausibilistic trace are shown in Figure 3 for $C = \{[a, b) \subseteq [0, 5)\}$. Five subset measurements are made yielding the measurement record

$$\vec{B} = \langle [3.5, 4), [3,4), [1,2), [1,2), [2, 2.5) \rangle.$$

After eliminating duplicates from $\vec{B}$, then the set of observed intervals $\mathcal{F}^E$ with $N = 4 < M = 5$ and random interval $\mathcal{A}^E$ are

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

$$\mathcal{A}^E = \{\langle [3.5, 4), .2\rangle, \langle [3,4), .2\rangle, \langle [1,2), .4\rangle, \langle [2, 2.5), .2\rangle\}.$$

$\mathcal{F}^E$ is inconsistent, and the support $U(\mathcal{A}^E) = [1, 2.5) \cup [3, 4)$ is not connected. The components of $\mathcal{F}^E$ are

$$\vec{E} = \langle 3.5, 3, 1, 2, 4, 4, 2, 2.5\rangle, \quad \vec{E}^l = \langle 3.5, 3, 1, 2\rangle, \quad \vec{E}^r = \langle 4, 4, 2, 2.5\rangle,$$

$$\hat{E} = \langle 1, 2, 3, 3.5, 2, 2.5, 4, 4\rangle, \quad E = \{1, 2, 2.5, 3, 3.5, 4\}, \quad E^l = \{1, 2, 3, 3.5\},$$

$$E^r = \{2, 2.5, 4\}.$$

with $Q = 6$, $Q^l = 4$, and $Q^r = 3$. The trace is the step function on the right of Figure 3 with

$$\rho([1, 2)) = .4, \quad \rho([2, 2.5)) = .2, \quad \rho([3, 3.5)) = .2, \quad \rho([3.5, 4)) = .4$$

and $\rho(x) = 0$ elsewhere.

Consistent Random Intervals

The following are some important properties of consistent random intervals.

THEOREM 42. (CONSISTENT ENDPOINTS) $\mathcal{A}$ is consistent iff

$$\max_j l_j = l_{(N)} < r_{(N)} = \min_j r_j, \tag{43}$$

with $\mathbf{C}(\mathcal{A}) = [l_{(N)}, r_{(N)})$.

Proof

Case 1: Assume $\mathcal{A}$ is consistent with core $\mathbf{C} = [l, r) := \mathbf{C}(\mathcal{A})$ and focal set $\mathcal{F} := \mathcal{F}(\mathcal{A})$, recalling that $l < r$. Fix an $x \in \mathbf{C}$. Since $\forall A_j \in \mathcal{F}$, $\mathbf{C} \subseteq A_j$, therefore $\forall A_j \in \mathcal{F}$, $x \in A_j$. So $\forall A_j \in \mathcal{F}$, $l_j \leq x$, and in particular $l_{(N)} = \max l_j \leq x$. For the right endpoints, $\forall A_j \in \mathcal{F}$, $x < r_j$, and in particular $x < r_{(N)} = \min r_j$. Thus $\forall x \in \mathbf{C}, l_{(N)} \leq x < r_{(N)}$, so that (43) holds, and $\mathbf{C} = [l, r) \subseteq [l_{(N)}, r_{(N)})$. If $\mathbf{C} \subset [l_{(N)}, r_{(N)})$, then either $\exists x \in [l_{(N)}, lc$ or $\exists x \in [r, r_{(N)})$. In either case, $\forall A_j \in \mathcal{F}$, $x \in A_j$, $x \notin \mathbf{C}$, which contradicts the consistency of $\mathcal{A}$. Thus $\mathbf{C} = [l, r) = [l_{(N)}, r_{(N)})$.

Case 2: Assume $l_{(N)} < r_{(N)}$. Consider an arbitrary $x \in [l_{(N)}, r_{(N)})$. Since $l_{(N)}$ is the rightmost left endpoint, therefore $\forall A_j$, $x \geq l_{(j)}$. Similarly, $\forall A_j \in \mathcal{F}$, $x < r_{(j)}$. Therefore for all such $x \in [l_{(N)}, r_{(N)})$, $\forall A_j$, $x \in A_j$, so that $\mathbf{C}(\mathcal{F}) = [l_{(N)}, r_{(N)}) \neq \emptyset$ and $\mathcal{A}$ is consistent by Definition (8). ∎

COROLLARY 44. (ENDPOINT ORDERING) $\mathcal{A}$ is consistent iff there is a joint linear order on $\hat{E}$

$$l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(N)} \leq r_{(N)} \leq r_{(N-1)} \leq \cdots \leq r_{(1)}.$$

Proof Trivial from Definition (37) and the consistent endpoint conditions of Theorem (42).

∎

LEMMA 45. (CONSISTENT RANDOM INTERVAL MONOTONICITY) If $\mathcal{A}$ is consistent then $\rho_{\mathcal{A}}$ is monotone nondecreasing from $-\infty$ to $\mathbf{C}(\mathcal{A})$ and monotone nonincreasing from $\mathbf{C}(\mathcal{A})$ to ∞.

Proof Denote $\rho := \rho_{\mathcal{A}}$. The proof will be carried out for $x \in [-\infty, r_{(N)})$. The remaining argument follows analogously for $x \in [l_{(N)}, \infty]$. Recall that endpoint ordering of Corollary (44) carries ove r to the $e^l_{k^l}$ and $e^r_{k^r}$.

1. For $x \in [-\infty, e^l_1)$, clearly $\nexists A_j$, $x \in A_j$, so $\rho(x) = 0$.
2. Let $I := [e^l_1, e^l_{Q^l})$. For $x \in I$, from the endpoint ordering of Corollary (44) $\nexists e_0 \in E^r$, $e_0 \in I$. Therefore from Theorem (41), cases 3 and 4, $\forall x, x' \in I, x < x' \rightarrow 0 \leq \rho(x) \leq \rho(x')$.
3. Finally, fix $x \in [e^l_{Q^l}, e^r_{Q^r})$. From Theorem (42), $x \in \mathbf{C}(\rho)$, so that $\rho(x) = 1$, and $\forall x' \in [-\infty, e^l_{Q^l})$, $\rho(x') \leq \rho(x)$.

For random intervals there is a form of Lemma (10) showing the equivalence of trace normalization and consistency.

∎

COROLLARY 46. (RANDOM INTERVAL NORMALIZATION AND CONSISTENCY) $\mathcal{A}$ is consistent iff

$$\sup_{x \in \mathbb{R}} \rho_{\mathcal{A}}(x) = 1. \tag{47}$$

Proof **Case 1:** If $\mathcal{A}$ is consistent, then (47) follows by the same argument as in Case 1 of the proof of Lemma (10). **Case 2:** Assume a random interval $\mathcal{A}$ with trace $\rho := \rho_{\mathcal{A}}$. From Theorem (41), we know that ρ is piecewise constant, with

$$\forall 1 \leq k \leq Q-1, \quad \forall x \in G_k, \quad \rho(x) = \lim_{y \to e_k^+} \rho(y) = \lim_{y \to e_{k+1}^-} \rho(y).$$

So for (47) to hold, there must be a G_k, $\rho(G_k) = 1$. The argument then follows as in Case 2 of the proof of Lemma (10).

Possibilistic Histograms

We now move to describe possibilistic histograms proper as possibility distributions derived from consistent empirical random intervals. ■

DEFINITION 48. (POSSIBILISTIC HISTOGRAM) Given a consistent empirical random interval $\mathcal{A}^E$, let the possibility distribution $\pi^E := \rho_{\mathcal{A}^E}$ derived according to the plausibility assignment formula (13) be called a **possibilistic histogram**.

Possibilistic histograms are similar to ordinary (stochastic) histograms, but resulting from overlapping interval observations, and thus governed by the mathematics of random sets.

COROLLARY 49. (POSSIBILISTIC HISTOGRAM FORM) If π^E is a possibilistic histogram, then

$$\mathrm{C}(\pi^E) = \left[e^l_{Q^l}, e^r_{Q^r}\right), \quad \mathrm{U}(\pi^E) = \left[e^l_1, e^r_1\right), \quad \pi^E\left(\left[-\infty, e^l_1\right)\right) = \pi^E([e^r_1, \infty]) = 0.$$

Proof Follows simply from the proof of Lemma (45) and the fact that $U(\pi^E)$ is connected, from Corollary (44).

Figure 4 shows an example of a possibilistic histogram for $C = \{[a, b) \subseteq [0, 5)\}$.

Possibilistic Histograms as Fuzzy Intervals

As stochastic histograms are naturally probability distributions, so possibilistic histograms are natural representations of possibility distributions. Since possibility theory is a weak representational form for uncertainty [Joslyn, 1993c], it is appropriate that they produce meaningful forms of possibility distributions even given very few observations. In particular, all possibilistic histograms are fuzzy intervals.

THEOREM 50. If a finite random interval $\mathcal{A}$ is consistent then its trace $\rho_{\mathcal{A}}$ is convex.

Proof Assume a consistent finite random interval $\mathcal{A}$ with trace $\rho := \rho_{\mathcal{A}}$. From Corollary (46), ρ is normal. Convexity follows from the following three cases, which themselves follow from the Lemma *(45). Let* $x, y, z \in \mathbb{R}$, $x \leq y \leq z$.

1. If $x \leq y \leq e^r_{Q^r}$ then $\pi(x) \wedge \pi(y) = \pi(x) \leq \pi(z)$.
2. If $e^l_{Q^l} \leq x \leq y$ then $\pi(x) \wedge \pi(y) = \pi(y) \leq \pi(z)$.

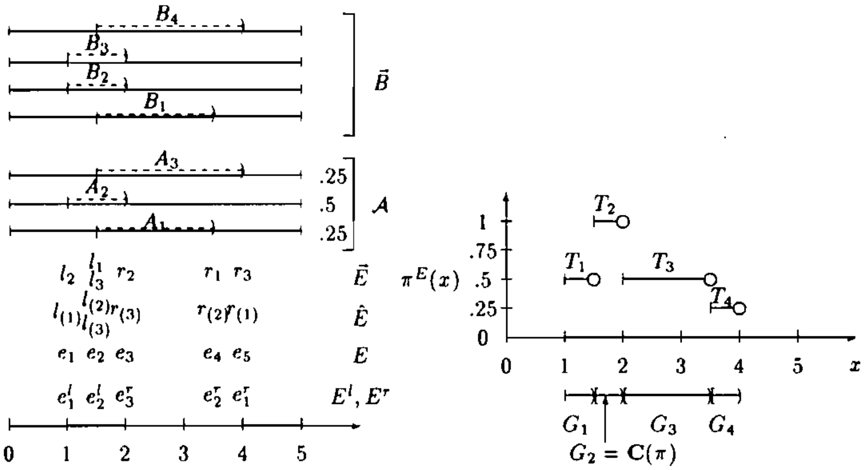

Figure 4 (Left) (top) A measurement record. (middle)$\mathcal{A}^E$ (bottom) Components of π^E. (Right) Possibilistic histogram π^E with more components.

3. If $x \le e^l_{Q^l} \le e^r_{Q^r} \le y$ then: if $x \le z \le e^r_{Q^r}$, then $\pi(x) \le \pi(z)$; similarly, if $e^l_{Q^l} \le z \le y$, then $\pi(y) \le \pi(z)$. Therefore $\pi(z) \ge \pi(x) \wedge \pi(y)$.

■

COROLLARY 51. If the trace $\rho_{\mathcal{A}}$ of a finite random interval $\mathcal{A}$ is normal, then it is convex.

Proof Let $\rho_{\mathcal{A}}$ be normal. Then from Corollary (46), $\mathcal{A}$ is consistent, and thus convex from Theorem (50).

We now arrive at the **central result** of this paper:

THEOREM 52. The trace $\rho_{\mathcal{A}^E}$ of an empirical finite random interval $\mathcal{A}^E$ is a possibilistic histogram π^E if an only if it is a fuzzy interval.

Proof **Case 1:** If $\rho_{\mathcal{A}^E}$ is a possibilistic histogram π^E, then from Definition (48) $\mathcal{A}^E$ is consistent, from Corollary (46) it is normal, from Corollary (51) it is convex, so from Definition (14) it is a fuzzy interval. **Case 2:** If $\rho_{\mathcal{A}^E}$ is a fuzzy interval, then from Definition (14) it is normal, from Corollary (46) $\mathcal{A}^E$ is consistent, and from Definition (48) $\rho_{\mathcal{A}^E}$ is a possibilistic histogram π^E.

COROLLARY 53. No possibilistic histogram π^E is a fuzzy number.

Proof From Theorem (52) we know that π^E is a fuzzy interval. For π^E to be a fuzzy number from Definition (15) there must be an $x \in \mathbb{R}$, $\mathbf{C}(\pi^E) = \{x\}$. But we know that this cannot be the case, since from Corollary (49) $C(\pi^E) = [e^l_{Q^l}, e^r_{Q^r})$. ■

CONTINUOUS APPROXIMATIONS

Possibilistic histograms play the role in possibility theory that ordinary histograms do in traditional statistics. As maximum likelihood and other estimation methods are used in statistics to generate continuous probability distributions which approximate the histogram, so it is desirable to develop continuous or smooth approximations to possibilistic histograms.

DEFINITION 54. (CONTINUOUS APPROXIMATION) Given a possibilistic histogram π^E, denote $\bar{\pi}$ as a possibility distribution which is continuous on $\mathbf{U}(\pi^E)$ and approximates π^E.

One of the most significant differences between possibilistic and stochastic histograms is that the former are collections of constant left-closed segments T_k of generally different lengths, not discrete points. Therefore, normal interpolation or approximation methods (such as curve-fitting or maximum-likelihood estimation) are not appropriate. Instead, a representative set of **candidate points** from the segments T_k should be selected, and then a continuous curve $\bar{\pi}$ fitted to *them.*

Candidate Points

First, characterize a possibilistic histogram π^E as

$$\pi^E = \Upsilon = \{T_k\} = \{\langle x, \pi^E(x)\rangle\} \subseteq \mathbf{U}(\pi^E) \times [0, 1].$$

Then it is necessary to characterize the candidate points from the possibilistic histogram.

DEFINITION 55. (POSSIBILISTIC HISTOGRAM CANDIDATE POINTS) Assume a possibilistic histogram $\pi^E = \Upsilon$. Then denote:

- The left and right endpoints of each of the T_k, $1 \le k \le Q - 1$:

$$\mathbf{t}^l_k := \langle e_k, \pi^E(e_k)\rangle, \quad \mathbf{t}^r_k := \langle e_{k+1}, \pi^E(e_k)\rangle.$$

- The midpoints of each of the T_k, $1 \le k \le Q - 1$:

$$\mathbf{h}_k := \left\langle \frac{e_k + e_{k+1}}{2}, \pi(e_k) \right\rangle.$$

- The midpoint of the core:

$$\mathbf{c} := \mathbf{h}_{Q'} = \left\langle \frac{l_{(N)} + r_{(N)}}{2}, 1 \right\rangle.$$

- The endpoints of the support at the axis:

$$\mathbf{l} := \langle l_{(1)}, 0\rangle, \quad \mathbf{r} := \langle r_{(1)}, 0\rangle.$$

- The set of all the interval mid- and end-points to which a continuous curve *may optionally* be fit: $\mathbf{K}' := \{\mathbf{t}^l_k, \mathbf{t}^r_k, \mathbf{h}_k\}$.
- The set of all these optional interval mid- and end-points to which a continuous curve *actually will* be fit: $\mathbf{K} \subseteq \mathbf{K}'$.
- Finally, the set of *all* the points to which the curve will be fit: $\mathbf{D} := \{\mathbf{c}, \mathbf{l}, \mathbf{r}\} \cup \mathbf{K} \subseteq \pi^E$.

The structure of $\mathbf{D}$ is then characterized by the following principle:

PRINCIPLE 56. (CANDIDATE POINT SELECTION) $\mathbf{K}$ may be any subset of $\mathbf{K}'$ such that $\forall x \in \mathbf{U}(\pi^E)$, there is at most one point in $\mathbf{K}$ for which x is the ordinate.

Note that $\mathbf{K} = \varnothing$ is allowed.

Both Definition (55) and Principle (56) are justified by the following argument:

1. Possibilistic normalization requires at least one point from the core to be a candidate. **c** is the only natural single point from the core, and so its requirement serves as the least restrictive normalization requirement.
2. For $\bar{\pi}$ to be zero outside the support $\mathbf{U}(\pi^E)$ and for $\bar{\pi}$ to be continuous, $\bar{\pi}$ should drop to the axis through the points **l** and **r**.
3. The above two criteria are the only *necessary* conditions to construct a continuous possibility distribution with support $\mathbf{U}(\pi^E)$. Therefore $\{\mathbf{c}, \mathbf{l}, \mathbf{r}\} \subseteq \mathbf{D}$, but **K** may be empty.
4. For each interval T_k, the naturally identifiable points, which are also consistent with the ordinal nature of possibilistic information, are $\mathbf{t}^l_k$, $\mathbf{t}^r_k$, and $\mathbf{h}_k$. Therefore they *may* be included in **K**.
5. The final requirement in Principle (56) is simply a statement that $\bar{\pi}$ must be a function, so that $\forall x \in \mathbf{U}(\bar{\pi})$, $\exists! \bar{\pi}(x)$. For example, this would preclude, for a fixed k, including both the right limit of T_k and the left endpoint of T_{k+1}, which are equal in x but differ in $\pi^E(x)$.

An Example

Consider the example in Figure 5. The left side shows two interval observations in dashed lines below the axis, each of which is observed once. The components of the T_k with $N = M = 2$, $Q = 3$, and $\mathbf{c} = \mathbf{h}_2$ are also shown. $\mathbf{t}^l_1$ and $\mathbf{t}^r_3$ are excluded from **K** due to conflicts with **l** and **r**, leaving a candidate set

$$\mathbf{K}' = \{\mathbf{h}_1, \mathbf{t}^r_1, \mathbf{t}^l_2, \mathbf{t}^r_2, \mathbf{t}^l_3, \mathbf{h}_3\}.$$

Any subset $\mathbf{K} \subseteq \mathbf{K}'$ (including the empty set) can be chosen as long as it does not contain either set of conflicts $\{\mathbf{t}^r_1, \mathbf{t}^l_2\}$ or $\{\mathbf{t}^r_2, \mathbf{t}^l_3\}$.

Piecewise Linear Approximations

Once a set of points is selected, a variety of curve-fitting methods are available to determine $\bar{\pi}$. The simplest and most direct is to connect them with line segments, producing a piecewise linear, continuous distribution. Three of these are shown on the right of Figure 5 for the sets

$$\mathbf{K} = \{\mathbf{h}_1, \mathbf{t}^l_2, , \mathbf{t}^r_2, \mathbf{h}_3\}, \quad \mathbf{K} = \emptyset, \quad \mathbf{K} = \{\mathbf{t}^r_1, \mathbf{t}^l_3\},$$

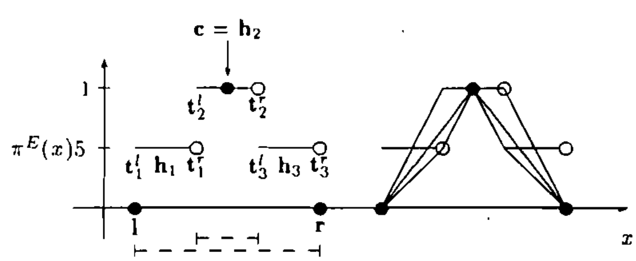

Figure 5 (Left) A simple possibilistic histogram with its candidate points. (Right) Three example piecewise linear continuous approximations.

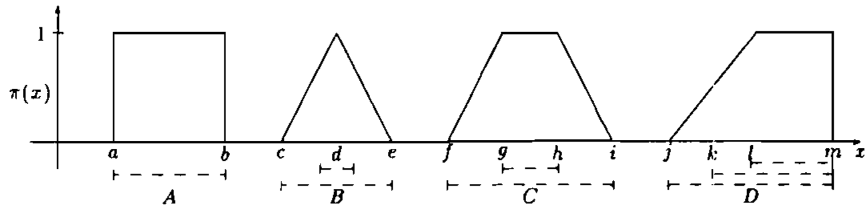

Figure 6 Typical fuzzy intervals and numbers used in applications.

moving from the outside to the inside respectively. Alternatively, nonlinear regression or spline methods can be used to fit the selected points to one of the exponential or quadratic forms which are commonly used for fuzzy numbers [Dubois and Prade, 1978; Tanaka and Ishibuchi, 1993].

An advantage of the line-segment method is that even given very few observations, $\bar{\pi}$ has the same form as the fuzzy intervals and numbers typically used in fuzzy systems applications. Some of these are shown in Figure 6, with some example observed intervals below them which could give rise to them. Case A is a square distribution produced by a single crisp interval $[a, b]$; B is the triangular form, produced in all cases when $d = \mathbf{c}$ and $\mathbf{K} = \varnothing$; C is the outermost case of Figure 5 for the observations $[f, i]$, $[g, h]$.

In case D it is also common for π to extend to the right by letting $m \rightarrow \infty$, so that $\forall x \geq l$, $\pi(x) = 1$. Either condition can result when point observations j, k, l are interpreted either as distances from a fixed m (perhaps an upper bound), or as magnitudes in relation to one or the other infinities. In this last case, π is simply equivalent to a cumulative probability distribution; but this approach is in keeping with the ordinal possibilistic concepts of capacity, distance, and similarity.

ACKNOWLEDGEMENTS

Sven Aerts and Gert de Cooman were very helpful with suggestions about the physical realization of consistent and consonant measuring devices. Discussion with Scott Ferson suggested some of the questions concerning the linkage of consistency of random intervals and convexity of their distributions. I would also like to thank Bo Yuan for his help and patience in editing this fine volume, and of course George Klir for all his help and support through the years.

REFERENCES

Artstein, Z. and R. A. Vitale [1975], "Strong Law of Large Numbers for Random Compact Sets." *Annals of Probability*, **3**, pp. 879–882.

Barone, Joseph M., Dimitar P. Filev, and Ronald R. Yager [1995], "Mountain Method-Based Fuzzy Clustering." *Int. J. General Systems*, **23**, pp. 281–305.

de Cooman, Gert [1995], "The Formal Analogy Between Possibility and Probability Theory." In: *Foundations and Applications of Possibility Theory*, edited by G. de Cooman, D. Ruan and E. E. Kerre, World Scientific, Singapore, pp. 71–87.

de Cooman, Gert [1997a], "Possibility Theory I: Measure- and Integral-Theoretic Groundwork." *Int. J. General Systems*, **25**(4), pp. 291–323.

de Cooman, Gert [1997b], "Possibility Theory II: Conditional Possibility." *Int. J. General Systems*, **25**(4), pp. 325–351.

de Cooman, Gert [1997c], "Possibility Theory III: Possibilistic Independence." *Int. J. General Systems*, **25**(4) pp. 353–371.

de Cooman, Gert, D. Ruan and E. E. Kerre, eds [1995], *Foundations and Applications of Possibility Theory*. World Scientific, Singapore.

David, Herbert A. [1981], *Order Statistics*. Wiley, New York.

Deely, John (1990) *Basics of Semiotics*. Indiana UP, Bloomington IN.

Degroot, Morris H. and William F. Eddy [1983], "Set-Valued Parameters and Set-Valued Statistics." In: *Recent Advances in Statistics*, Academic Press, pp. 175–195.

Dempster, A. P. [1967], "Upper and Lower Probabilities Induced by a Multivalued Mapping." *Annals of Mathematical Statistics*, **38**, pp. 325–339.

Dubois, Didier and Henri Prade [1978], "Operations on Fuzzy Numbers." *Int. J. Systems Science*, **9**, pp. 613–626.

Dubois, Didier and Henri Prade [1982], "On Several Representations of an Uncertain Body of Evidence." *Fuzzy Information and Decision Processes*, pp. 167–181.

Dubois, Didier and Henri Prade [1987], "Fuzzy Numbers: An Overview." In: *Analysis of Fuzzy Information*, CRC Press, Boca Raton, v. **1**, pp. 3–39.

Dubois, Didier and Henri Prade [1989], "Fuzzy Sets, Probability, and Measurement." *European J. of Operational Research*, **40**(2), pp. 135–154.

Dubois, Didier and Henri Prade [1990], "Consonant Approximations of Belief Functions." *Int. J. Approximate Reasoning*, **4**, pp. 419–449.

Dubois, Didier and Henri Prade [1992], "Evidence, Knowledge and Belief Functions." *Int. J. Approximate Reasoning*, **6**(3), pp. 295–320.

Eco, Umberto [1986], *Semiotics and the Philosophy of Language*. Indiana U Press, Bloomfield.

Ferson, Scott, Lev Ginzburg, and Resit Akcakaya [1996], "Whereof One Cannot Speak: When Input Distributions are Unknown." *Risk Analysis*, in press.

Fung, Robert M. and Chee Yee Chong [1986], "Metaprobability and the Dempster-Shafer in Evidential Reasoning." In: *Uncertainty in Artificial Intelligence*, edited by L. N. Kanal and J. F. Lemmer, North-Holland, Amsterdam, pp. 295–302.

Gil, M. A. [1992], "A Note on the Connection Between Fuzzy Numbers and Random Intervals." *Statistics and Probability Letters*, **13**(4), pp. 311–319.

Goodman, I. R. [1994], "New Characterization of Fuzzy Logic Operators Producing Homomorphic-Like Relations with One-Point Coverages of Random Sets." In: *Advances in Fuzzy Theory and Technology*, edited by P. P. Wang, Bookwrights Press, Raleigh NC, pp. 133–159.

Halmos, P. R. [1950], *Measure Theory*. van Nostrand, New York.

Janssen, Hugo, J., Gert de Cooman, and E. E. Kerre [1996], "First Results for a Mathematical Theory of Possibilistic Processes." In: *Proc. 1996 European Meeting on Cybernetics and Systems Research*, edited by R. Trappl, pp. 341–346.

Joslyn, Cliff [1992], "Possibilistic Measurement and Set Statistics." In: *Proc. NAFIPS 1992*, v. **2**, Puerto Vallerta, pp. 458–467.

Joslyn, Cliff [1993a], "Empirical Possibility and Minimal Information Distortion." In: *Fuzzy Logic: State of the Art*, edited by R. Lowen and M. Roubens, Kluwer, Dordrecht, pp. 143–152.

Joslyn, Cliff [1993b], "Some New Results on Possibilistic Measurement." In: *Proc. NAFIPS 1993*, Allentown PA, pp. 227–231. To be reprinted in: *Fuzziness: An Introduction to Theory and Applications*, edited by M. Mcallister, SIAM Press.

Joslyn, Cliff [1993c], "Possibilistic Semantics and Measurement Methods in Complex Systems." In: *Proc. 2nd International Symposium on Uncertainty Modeling and Analysis*, editted by Bilal Ayyub, IEEE Computer Society Press, pp. 208–215.

Joslyn, Cliff [1994a], *Possibilistic Processes for Complex Systems Modeling*. UMI Disseration Services, Ann Arbor MI (PhD dissertation, SUNY-Binghamton).

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

Joslyn, Cliff [1994c], "On Possibilistic Automata." In: *Computer Aided Systems Theory—EUROCAST '93*, edited by F. Pichler and R. Moreno-Diáz, Springer-Verlag, Berlin, pp. 231–242.

Joslyn, Cliff [1995a], "In Support of an Independent Possibility Theory." In: *Foundations and Applications of Possibility Theory*, edited by G. de Cooman, D. Ruan and E. E. Kerre, World Scientific, Singapore, pp. 152–164.

Joslyn, Cliff [1995b], "Strong Probabilistic Compatibility of Possibilistic Histograms." In: *Proc. 1995 Int. Symposium on Uncertainty Modeling and Analysis*, edited by Bilal Ayyub, IEEE Computer Society Press, Los Alamitos CA, pp. 383, A17–22.

Joslyn, Cliff [1995c], "Semantic Control Systems." *World Futures*, **45**, pp. 87–123.

Kendall, D. G. [1974], "Foundations of a Theory of Random Sets." In: *Stochastic Geometry*, edited by E. F. Harding and D. G. Kendall, Wiley, New York, pp. 322–376.

Klir, George [1991], "Generalized Information Theory." *Fuzzy Sets and Systems*, **40**, pp. 127–142.
Klir, George [1993], "Developments in Uncertainty Based Information." In: *Advances in Computers*, v. **36**, edited by M. Yovitz, Academic Press, pp. 255–332.
Klir, George and Behzad Parviz [1992], "Possibility-Probability Conversions: An Empirical Study." In: *Progress in Cybernetics and Systems*, edited by R. Trappl, Hemisphere, New York, pp. 19–26.
Klir, George and Bo Yuan [1995], *Fuzzy Sets and Fuzzy Logic*. Prentice-Hall, Upper Saddle River, N.J.
Krishnapuram, Raghu and James Keller [1993], "Possibilistic Approach to Clustering." *IEEE Trans. on Fuzzy Systems*, **1**(2), pp. 98–110.
Pawlak, Zdzislaw [1991], *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer, Boston.
Rosen, Robert [1985], *Anticipatory Systems*. Pergamon, Oxford.
Shafer, Glen [1976], *Mathematical Theory of Evidence*. Princeton University Press, Princeton.
Tanaka, Hideo and Hisao Ishibuchi [1993], "Evidence Theory of Exponential Possibility Distribution." *Int. J. Approximate Reasoning*, **8**, pp. 123–140.
Walley, P. [1990], *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York.
Wang, P. Z. and X. H. Liu [1984], "Set-Valued Statistics." *J. Engineering Mathematics*, **1**(1), pp. 43–54.
Wang, P. Z., X. H. Liu and E. Sanchez [1986], "Set-Valued Statistics and its Application to Earthquake Engineering." *Fuzzy Sets and Systems*, **18**, pp. 347–356.
Wang, Zhenyuan and George Klir [1992], *Fuzzy Measure Theory*. Plenum Press, New York.

Cliff Joslyn received his B.A. in Cognitive Science and Mathematics from Oberlin College in 1985, with High Honors in Systems Theory, and his M.S. in 1989 and Ph.D. in 1994, both in Systems Science from the State University of New York at Binghamton. In his dissertation he developed the semantic and mathematical basis for possibilistic measurement and modeling. He is currently a Postdoctoral Scientist at the Los Alamos National Laboratory, where he pursues research on data mining and data fusion using possibilistic and fuzzy systems theory. His research interests include Systems Science, General Information Theory (GIT), qualitative modeling, Soft Methods for Discrete EVent Systems (DEVS) modeling, Computer-Aided Systems Theory (CAST), biological semiotics, cybernetic philosophy, and distributed hypertext development environments.