Joslyn, Cliff A: (1993) "Some New Results on Possibilistic Measurement", in: Proc. 1993 Conf. of the North American Fuzzy Information Processing Society (NAFIPS 1993), pp. 227-231, Allentown PA

Some New Results on Possibilistic Measurement*

Cliff Joslyn†

Abstract

Further results on possibilistic measurement [5, 8, 9] are presented, including the introduction of possibilistic histograms, their interpretation as fuzzy numbers, and their continuous approximations.

1 Possibilistic Measurement

Joslyn has presented a measurement method for possibility distributions [5, 8, 9]. The procedure is based on the observations of possibly non-disjoint intervals. From these **set statistics** an **empirical random set** can be derived. Under reasonable consistency requirements, its one-point coverage function is a possibility distribution, from which a consonant (possibilistic) random set can in turn be derived.

Given a finite universe $\Omega := \{\omega_i\}, 1 \le i \le n$, the function $m\colon 2^\Omega \mapsto [0,1]$ is an **evidence function** (otherwise known as a **basic probability assignment**) when $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. Denote a random set generated from an evidence function as $\mathcal{S} := \{\langle A_j, m_j \rangle : m_j > 0\}$, where $\langle \cdot \rangle$ is a vector, $A_j \subseteq \Omega, m_j := m(A_j)$, and $1 \le j \le N := |\mathcal{S}| \le 2^n - 1$. Denote the **focal set** as $\mathcal{F} := \{A_j : m_j > 0\}$ with **core** $\mathbf{C}(\mathcal{F}) := \bigcap_{\mathcal{F}} A_j$ and **support** $\mathbf{S}(\mathcal{F}) := \bigcup_{\mathcal{F}} A_j$.

The plausibility measure on $\forall A \subseteq \Omega$ is $\mathrm{Pl}(A) := \sum_{A_j \cap A \ne \emptyset} m_j$. The **plausibility assignment** (otherwise known as the **one-point coverage function**) of $\mathcal{S}$ is

$$\vec{\mathrm{Pl}} = \langle \mathrm{Pl}_i \rangle := \langle \mathrm{Pl}(\{\omega_i\}) \rangle, \qquad \mathrm{Pl}_i := \sum_{A_j \ni \omega_i} m_j.$$

$\vec{\mathrm{Pl}}$ is a fuzzy set that can be mapped to an equivalence class of random sets [10].

When $\forall A_j \in \mathcal{F}, |A_j| = 1$, then $\mathcal{S}$ is **specific**, and $\mathrm{Pr}(A) := \mathrm{Pl}(A)$ is an additive **probability measure** with **probability distribution** $\vec{p} = \langle p_i \rangle := \vec{\mathrm{Pl}}$ and additive normalization $\sum_i p_i = 1$ and operator $P(A) = \sum_{\omega_i \in A} p_i$. $\mathcal{S}$ is **consonant** ($\mathcal{F}$ 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**. As Pr is additive, so Π is **maximal**:

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

where $\vee$ is the maximum operator. Denoting $A_i := \{\omega_1, \omega_2, \ldots, \omega_i\}$, and assuming that $\mathcal{F}$ is complete (i.e. $\forall \omega_i \in \Omega, \exists A_i$), then $\vec{\pi} = \langle \pi_i \rangle := \vec{\mathrm{Pl}}$ is a **possibility distribution** with: support $\mathbf{S}(\pi) := \{\omega\colon \pi(\omega) > 0\} = \mathbf{S}(\mathcal{F})$; core $\mathbf{C}(\pi) := \{\omega\colon \pi(\omega) = 1\} = \mathbf{C}(\mathcal{F}) \ne \emptyset$; maximal normalization $\bigvee_i \pi_i = 1$; and operator $\Pi(A) = \bigvee_{\omega_i \in A} \pi_i$.

However, it is not necessary that $\mathcal{S}$ be consonant for $\bigvee_i \pi_i = 1$. The weaker condition of **consistency** $\mathbf{C}(\mathcal{F}) \ne \emptyset$ is sufficient, and therefore this is all that is required for $\mathcal{S}$ to have a

*Prepared for *Fuzziness: An Introduction to Theory and Applications*, ed. M. Mcallister, SIAM. Previously appeared in *Proceedings of the 1993 Conference of the North America Fuzzy Information Processing Society*, pp. 227-231, Allentown PA.

†NRC Research Associate, Mail Code 522.3, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA, joslyn@kong.gsfc.nasa.gov, http://pespmc1.vub.ac.be/joslyn, (301) 286-5773.

possibility distribution. If $\mathcal{S}$ is not consistent, then there are well-justified possibilistic normalization methods [6, 12].

To derive a possibility distribution from an empirical source, it is necessary to observe subsets $B_s \subseteq \Omega, 1 \leq s \leq M$ denoted as a vector $\vec{B} := \langle B_s \rangle$. The set of observed subsets produced by eliminating duplicates in $\vec{B}$ is an empirically derived focal set $\mathcal{F}^E := \{A_j\}, N \leq M$. Denoting the number of observed subsets A_j as C_j, then the set-frequency function is

$$m^E \colon \mathcal{F}^E \mapsto [0,1], \qquad m^E(A_j) = \frac{C_j}{\sum_{A_j \in \mathcal{F}^E} C_j} = C_j/M.$$

m^E in turn generates an empirically derived random set $\mathcal{S}^E$. If $\mathcal{F}^E$ is a disjoint class, then $\mathcal{S}^E$ generates a probability distribution on an equivalence class on $\mathbf{S}(\mathcal{F})$. But if $\mathcal{S}^E$ is consistent, or is a consistent approximation, then the empirical possibility distribution is

$$\pi(\omega) = \frac{\sum_{A_j \ni \omega} C_j}{M}. \tag{1}$$

In [6] a variety of methods to gather set statistics are outlined, including: 1) ensembles of heterogeneous measuring devices, changing the backdrop of measurement from multiple time trials on a single instrument to multiple *instrument* trials at a *single* time; and 2) intervals derived from order statistics of observed (singleton) point data.

2 Possibilistic Histograms

Possibility distributions derived according to (1) can be properly described as **possibilistic histograms**, similar to ordinary (stochastic) histograms, but derived from possibly overlapping interval observations, and thus governed by the mathematics of random sets. In the sequel it will be assumed that $\mathcal{S}^E$ is consistent, either naturally or as the result of a normalization method, and thus π from (1) is a possibility distribution.

Let $\Omega = \mathbb{R}$, and each observation $A_j \in \mathcal{F}^E$ be a closed interval denoted by its endpoints $A_j := [l_j, r_j]$. Let $l_{(j)}$ and $r_{(j)}$ be the order and "reverse order" statistics [1] of the left and right endpoints, so that

$$l_{(1)} \leq l_{(2)} \leq \ldots \leq l_{(N)}, \qquad r_{(N)} \leq r_{(N-1)} \leq \ldots \leq r_{(1)}. \tag{2}$$

are permutations of the l_j, r_j. Denote the multisets of endpoints and ordered endpoints as the vectors

$$\vec{E} := \langle l_1, l_2, \ldots, l_N, r_1, r_2, \ldots, r_N \rangle, \qquad \hat{E} := \left\langle l_{(1)}, l_{(2)}, \ldots, l_{(N)}, r_{(N)}, r_{(N-1)}, \ldots, r_{(1)} \right\rangle.$$

Consistency requires that

$$\max_j l_j = l_{(N)} \leq r_{(N)} = \min_j r_j, \tag{3}$$

so that $\mathbf{C}(\pi) = [l_{(N)}, r_{(N)}]$. If $l_{(N)} = r_{(N)}$ then π has a point core. The joint linear order on $\hat{E}$ is then

$$l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(N)} \leq r_{(N)} \leq r_{(N-1)} \leq \cdots \leq l_{(1)}. \tag{4}$$

The inequalities in (2) 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 first a single observation $A_1 := [a, b]$. When $a = b$ then A_1 is a point observation. When a second observation $A_2 := [c, d]$ is made, then there are four possibilities:

$$c = d \in \{a, b\}, \qquad c \in \{a, b\}, d \notin \{a, b\}, \qquad c \notin \{a, b\}, d \in \{a, b\}, \qquad c \notin \{a, b\}, d \notin \{a, b\}.$$

As distinct, consistent, observed intervals are added, in one limit all the l_j, r_j are distinct; in the other they all share only a common point core $r_{(N)} = l_{(N)}$. Let $E := \{e_k\}, 1 \leq k \leq P := |E|$ be the set of endpoints with duplicates omitted from $\vec{E}$. Then in general $N + 1 \leq P \leq 2N$.

Each of the e_k is equal to at least one of the (left or right) endpoints. From (4), the e_k naturally divide into the two groups mapping to left and right endpoints. Therefore denote $E = E^l \cup E^r$, where $E^l := \{e^l_{k^l}\}, E^r := \{e^r_{k^r}\}$ are the left- and right-endpoints respectively, ordered as in (4), where

$$P^l := |E^l|, \quad 1 \leq k^l \leq P^l, \qquad P^r := |E^r|, \quad P^r \geq k^r \geq 1, \qquad P^l + P^r = P.$$

π is completely determined by the coordinates $\langle\, e_k, \pi(e_k) \,\rangle$. First, $\mathbf{S}(\pi) = [e^l_1, e^r_1]$ and $\mathbf{C}(\pi) = [e^l_{P^l}, e^r_{P^r}]$. In general π is piecewise constant. Each e_k marks a discrete jump either up to $\pi(e_k)$ or down to $\pi(e_k + 1)$, depending on whether $e_k \in E^r$ or $e_k \in E^l$. For an (open or closed) interval $I \subseteq \mathbb{R}$ and $y \in [0, 1]$, let $\pi(I) = y$ denote that $\forall x \in I, \pi(x) = y$. Then

$$\pi([-\infty, e^l_1)) = \pi((e^r_1, \infty]) = 0, \quad \pi([e^l_1, e^l_2)) = \pi(e^l_1), \quad \ldots, \quad \pi([e^l_{P^l-1}, e^l_{P^l}) = \pi(e^l_{P^l-1}),$$

$$\pi([e^l_{P^l}, e^r_{P^r}]) = 1, \quad \pi((e^r_{P^r}, e^r_{P^r-1}]) = \pi(e^r_{P^r-1}), \quad \ldots, \quad \pi((e^r_2, e^r_1]) = \pi(e^r_1)$$

Letting $G_k, 1 \leq k \leq P - 1$ be the (appropriately half-open or closed) interval in $\mathbb{R}$ from e_k to e_{k+1}, then let $D_k, 1 \leq k \leq P - 1$ be the locus of points $\{\langle\, x, y \,\rangle : x \in G_k, y = \pi(x)\}$ which comprise the actual points of π.

2.1 Example

As an example, consider the vector of interval observations

$$\vec{B} = \langle\, [1.5, 3.5], [1, 2], [1, 2], [1.5, 4] \,\rangle\,,$$

so that $M = 4$ (see Fig. 1). Then $N = 3, l_{(N)} = 1.5$, and $r_{(N)} = 2$, and $\mathbf{C}(\pi) = [1.5, 2], \mathbf{S}(\pi) = [1, 4]$. Furthermore, $P = 5$ and $e_2 = 1.5$ maps to $l_1 = l_3 = l_{(2)} = l_{(3)} = e_2 = e^l_2$ with $P^l = 2$ and $P^r = 3$. The various sets and vectors are (the D_k and G_k are shown in the figure):

$$\mathcal{F}^E = \{[1, 2], [1.5, 3.5], [1.5, 4]\}, \qquad \vec{E} = \langle\, 1.5, 1, 1.5, 3.5, 2, 4 \,\rangle\,, \qquad \hat{E} = \langle\, 1, 1.5, 1.5, 2, 3.5, 4 \,\rangle\,,$$

$$E = \{1, 1.5, 2, 3.5, 4\}, \qquad E^l = \{1, 1.5\}, \qquad E^r = \{2, 3.5, 4\}.$$

2.2 Possibilistic Histograms as Fuzzy Numbers

Possibilistic histograms are natural representations of possibility distributions. Since possibility theory is a weak representational form for uncertainty [4, 7], it is appropriate that they produce meaningful forms of possibility distributions even given very few observations. In particular, possibilistic histograms are fuzzy intervals, and those with point cores are fuzzy numbers.

Lemma 1 *π is monotone increasing from $-\infty$ to $\mathbf{C}(\pi)$ and monotone decreasing from $\mathbf{C}(\pi)$ to ∞.*

Proof: Let $x \in \mathbb{R}$. The proof will be carried out for $x \in [-\infty, r_{(N)}]$; the remaining argument follows analogously for $x \in [l_{(N)}, \infty]$. First, recall that the ordering of (4) carries over to the e^l and

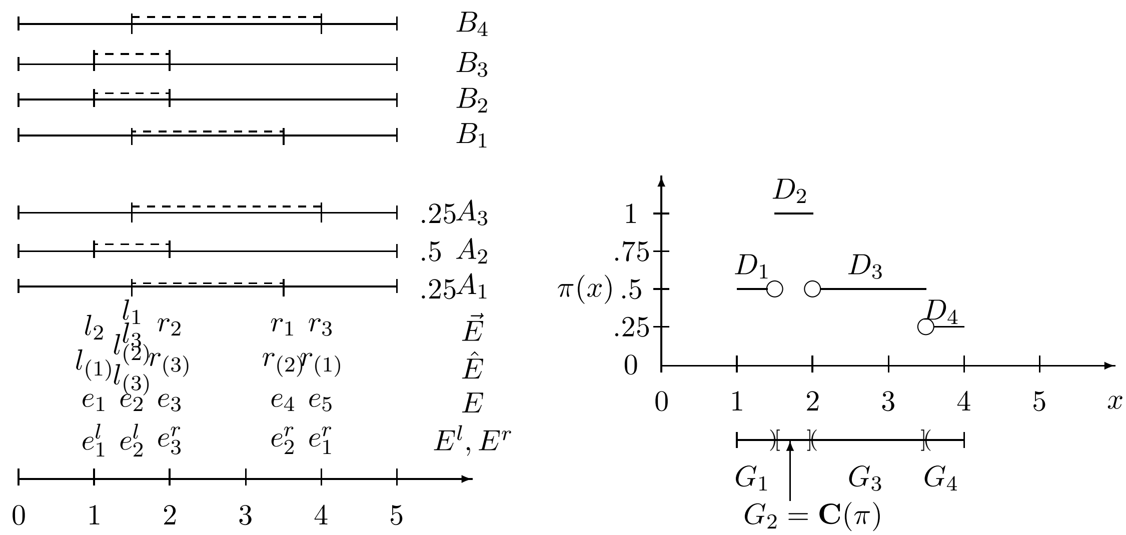

Figure 1: Four observed intervals, their statistics, and their possibilistic histogram.

e^r. When $x < e^l_1$ then $\pi(x) = 0$. Then, let $1 \le k \le P^l$ and let $x_k \in [e^l_k, e^l_{k+1})$, so that

$$\pi(x_k) = \pi(e^l_k) = \sum_{A_j \ni x_k} C_j/M \ = \sum_{A_j \supseteq [e^l_k, e^l_{k+1})} C_j/M.$$

From (3) and (4), $\forall e^l_{k^l}, e^r_{k^r}, e^l_{k^l} \le e^l_{k^l+1} \le e^l_{k^l+2} \le e^r_{k^r}$. Therefore

$$|\{A_j \colon A_j \supseteq [e^l_k, e^l_{k+1})\}| \quad \le \quad |\{A_j \colon A_j \supseteq [e^l_{k+1}, e^l_{k+2})\}|,$$

and so $\pi(x_k) \le \pi(x_{k+1}) \le 1$. Finally, when $x \in [e^l_{Pl}, e^r_{Pr}] = \mathbf{C}(\pi)$, then $\pi(x) = 1$. ■

Definition 1 (Fuzzy Interval) *[3] A fuzzy subset of the real numbers F is a fuzzy interval if 1)* Possibilistic Normalization: *$\mathbf{C}(F) \ne \emptyset$; and 2)* Convexity: *$\forall x, y \in \mathbb{R}, \forall z \in [x, y], \mu_F(z) \ge \mu_F(x) \wedge \mu_F(y)$.*

Definition 2 (Fuzzy Number) *[3] A fuzzy interval F is a fuzzy number if $\exists r \in \mathbb{R}, \mathbf{C}(F) = \{r\}$.*

Theorem 1 *If $\mathcal{F}^E$ is consistent, then π is a fuzzy interval.*

Proof: (1) Condition 1 of Def. 1 is satisfied by the consistency of $\mathcal{F}^E$. (2) For condition 2 of Def. 1, there are three cases, all of which follow from Lem. 1. a) If $x \le y \le e^r_{Pr}$ then $\pi(x) \wedge \pi(y) = \pi(x) \le \pi(z)$. b) If $e^l_{Pl} \le x \le y$ then $\pi(x) \wedge \pi(y) = \pi(y) \le \pi(z)$. c) If $x \le e^l_{Pl} \le e^r_{Pr} \le y$ then: if $x \le z \le e^r_{Pr}$, then $\pi(x) \le \pi(z)$; similarly, if $e^l_{Pl} \le z \le y$, then $\pi(y) \le \pi(z)$. Therefore $\pi(z) \ge \pi(x) \wedge \pi(y)$. ■

Corollary 1 *If $\exists r \in \mathbb{R}, \mathbf{C}(\mathcal{F}^E) = \{r\}$, then π is a fuzzy number.*

Proof: Obvious. ■

Note that: from condition 1 of Def. 1, fuzzy intervals and numbers are in fact possibility distributions; the instrument ensemble methods introduced in [5] typically produce fuzzy intervals, while the order statistical methods typically produce fuzzy numbers.

3 Continuous Approximations

Possibilistic histograms are to possibility theory as ordinary histograms are to traditional statistics. As maximum likelihood and other estimation methods are used in statistics to generate continuous approximations to histograms, so it is desirable to develop continuous or smooth approximations to possibilistic histograms.

One of the most significant differences between possibilistic and stochastic histograms is that the former are collections of intervals, not discrete points. We can proceed by selecting a set of points from these intervals to which a continuous curve will be fitted.

The following ideas suggest themselves:

- The midpoint of the core, $\mathbf{c} := \left\langle \frac{l_{(N)}+r_{(N)}}{2}, 1 \right\rangle$, should always be selected.
- To facilitate a smooth drop to the axis at the edge of the support, then if the points $\mathbf{l} := \left\langle l_{(1)}, 0 \right\rangle$ and $\mathbf{r} := \left\langle r_{(1)}, 0 \right\rangle$ do not equal $\mathbf{c}$, then they should always be selected.
- The endpoints of each D_k, denoted $\mathbf{d}_k^l := \langle e_k, \pi(e_k) \rangle$ and $\mathbf{d}_k^r := \langle e_{k+1}, \pi(e_k) \rangle$ should be candidates.
- The midpoints of each of the D_k, denoted $\mathbf{h}_k := \left\langle \frac{e_k+e_{k+1}}{2}, \pi(e_k) \right\rangle$ should be candidates.

Given a set of required and candidate points, the only other criterion is that only one point is selected for each $x \in \mathbf{S}(\pi)$. This would preclude, for example, including both the right limit of a D_k open on the right and the left limit of D_{k+1} closed on the left, which are equal in x but differ in $\pi(x)$.

Regrettably, space precludes a detailed analysis. Let the example in Fig. 2 suffice to illustrate the approach. The left side shows two observed intervals, in dashed lines below the axis, and the components of the D_k with $N = M = 2$ and $P = 3$. The set of required points is $\{\mathbf{c} = \mathbf{h}_2, \mathbf{l}, \mathbf{r}\}$. $\mathbf{d}_1^l$ and $\mathbf{d}_3^r$ are excluded due to conflicts with $\mathbf{l}$ and $\mathbf{r}$, leaving a candidate set $\{\mathbf{h}_1, \mathbf{d}_1^r, \mathbf{d}_2^l, \mathbf{d}_2^r, \mathbf{d}_3^l, \mathbf{h}_3\}$. Any subset $\mathbf{D}$ (including the empty set) can be chosen as long as it does not contain either $\{\mathbf{d}_1^r, \mathbf{d}_2^l\}$ or $\{\mathbf{d}_2^r, \mathbf{d}_3^l\}$.

Once a set of points is selected, a variety of curve-fitting methods are available. 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 Fig. 2 for the sets $\mathbf{D} = \{\mathbf{h}_1, \mathbf{d}_2^l, \mathbf{d}_2^r, \mathbf{h}_3\}, \emptyset, \{\mathbf{d}_1^r, \mathbf{d}_3^l\}$, moving from the outside to the inside. Alternatively, nonlinear regression or spline methods can be used to fit the selected points to an exponential or quadratic form, also commonly used for fuzzy numbers [2, 11, 13].

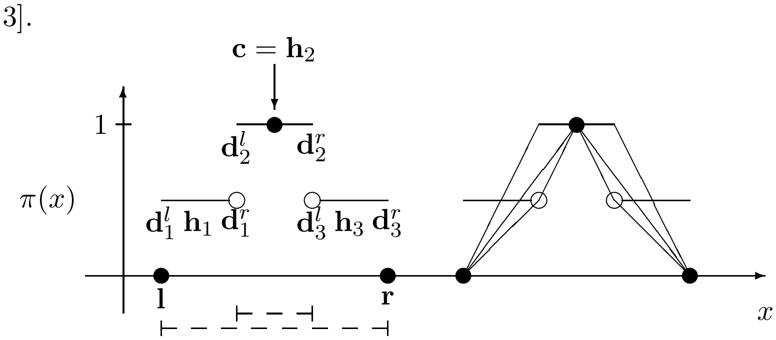

Figure 2: Example piecewise linear continuous possibility distributions.

An advantage of the line-segment method is that even given very few observations, the possibility distributions have the same form as those typically used in applications. Some of these are shown in Fig. 3, 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 $\mathbf{D} = \emptyset$ is selected; C is the outermost case of Fig. 2 for the observations $[f, i], [g, h]$.

In case D it is also common for π to extend to the right by letting $m \longrightarrow \infty$, 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 possibilistic semantics [7], which draws from the ordinal concepts of capacity, distance, and similarity.

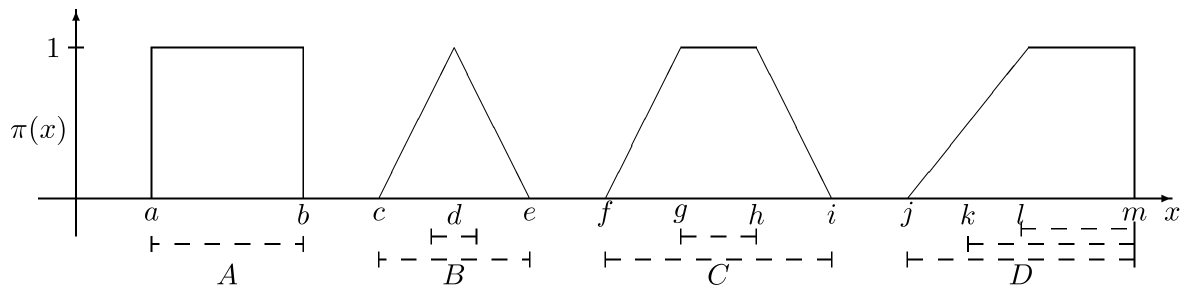

Figure 3: Typical fuzzy intervals and numbers used in applications.

References

[1] David, Herbert A: (1981) *Order Statistics*, Wiley, New York

[2] Dubois, Didier and Prade, Henri: (1978) "Operations on Fuzzy Numbers", *Int. J. Systems Science*, v. **9**, pp. 613-626

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

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

[5] Joslyn, Cliff: (1992) "Possibilistic Measurement and Set Statistics", in: *Proc. NAFIPS 1992*, v. **2**, pp. 458-467, Puerto Vallerta

[6] Joslyn, Cliff: (1993) "Empirical Possibility and Minimal Information Distortion", in: *Fuzzy Logic: State of the Art*, ed. R Lowen and M Roubens, pp. 143-152, Kluwer, Dordrecht

[7] Joslyn, Cliff: (1993) "Possibilistic Semantics and Measurement Methods in Complex Systems", in: *Proc. Second ISUMA Conference*, ed. Bilal Ayyub, pp. 208-215, IEEE Computer Society Press

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

[9] Joslyn, Cliff: (1996) "Measurement of Possibilistic Histograms from Interval Data", *Int. J. General Systems*, submitted

[10] Kampè de, Fériet J: (1982) "Interpretation of Membership Functions of Fuzzy Sets in Terms of Plausibility and Belief", in: *Fuzzy Information and Decision Processes*, ed. MM Gupta, pp. 93-98, North-Holland, Amsterdam

[11] Kreinovich, V; Quintana, C, and Reznik, L: (1992) "Gaussian Membership Functions Are Most Adequate in Representing Uncertainty in Measurements", in: *Proc. NAFIPS 1992*, v. **2**, pp. 618-624

[12] Ramer, Arthur and Puflea-Ramer, Rodica: (1993) "Uncertainty as the Basis of Possibility Conditioning", in: *Proc. Second ISUMA Conference*, ed. BM Ayyub, pp. 79-82, IEEE Computer Society Press

[13] Tanaka, Hideo and Ishibuchi, Hisao: (1993) "Evidence Theory of Exponential Possibility Distribution", *Int. J. Approximate Reasoning*, v. **8**, pp. 123-140