Joslyn, Cliff A: (1991) "Towards an Empirical Semantics of Possibility Through Maximum Uncertainty", in: Proc. 1991 Int. Fuzzy Systems Association World Conf. (IFSA 91), v. A, pp. 86-89

# Towards an Empirical Semantics of Possibility Through Maximum Uncertainty*

Cliff Joslyn

**Systems Science, SUNY–Binghamton**
**Binghamton, New York 13901, USA**

The mathematical syntax of Possibilistic Information Theory can be based on Zadeh's theory of fuzzy sets [19], the Dempster-Shafer theory of evidence [15], or the theory of random sets [14]. But the semantics for interpreting possibility measures and distributions is not as well developed. The traditional semantics of possibility is based on the use of "linguistic variables" or other means of deriving possibility measures from the opinions of people (for example, "experts" in some field). Similarly, the applications of possibility and fuzzy set theory are overwhelmingly in areas of "informational engineering" such as control theory, approximate reasoning, and decision support. Even in those attempts to apply fuzzy methods to the theory or modeling of natural, physical systems, data is predominantly collected on the basis of opinion, e.g. [8].

This situation contrasts sharply with traditional (probabilistic) information theory, which developed in close relation to the physics of many-body problems and thermodynamic systems [1] and natural communications [16]. It is a tacit assumption of most fuzzy researchers that while probabilistic randomness can be measured and applied in the physical world, possibilistic fuzziness is purely a result of human psychology, perception, and description. It is our interest to derive possibilisties on the basis of evidence. We will do so through the application of the Principle of Maximum Uncertainty to set-valued statistics.

## Mathematical Preliminaries

First, we have the standard evidence theory. For a finite universe $U = \{x_i\}$ with power set $2^U = \{A \subset U\}, m\colon 2^U \mapsto [0,1]$ is a function on the subsets of $U$ with $m(\emptyset) = 0$ and $\sum_{A \subset U} m(A) = 1$. Denote a random set as $\mathcal{S} = \{\langle A_j, m_j \rangle : m_j > 0\}$, where $\langle \cdot \rangle$ is a vector, $A_j \subset U, m_j = m(A_j)$, and $1 \leq j \leq |\mathcal{S}| \leq 2^{|U|} - 1$. We also have the focal set $\mathcal{F} = \{A_j : m_j > 0\}$. Then the dual belief and plausibility measures on an $A \in \mathcal{F}$ are:

$$\mathrm{Bel}(A) = \sum_{A_j \subset A} m_j = 1 - \mathrm{Pl}(\bar{A}) \tag{1}$$

$$\mathrm{Pl}(A) = \sum_{A_j \cap A \neq \emptyset} m_j = 1 - \mathrm{Bel}(\bar{A}). \tag{2}$$

We denote the "plausibility distribution" of a random set $\mathcal{S}$ as $\vec{\mathrm{Pl}} = \langle \mathrm{Pl}(\{x_i\}) \rangle = \langle \mathrm{Pl}_i \rangle$.

Klir and Ramer [12] identify two complementary uncertainty measures on random sets. The first is the *discord*:

$$\mathbf{D}(\mathcal{S}) = -\sum_j m_j \log_2 \left[ \sum_{k=1}^{|\mathcal{S}|} m_k \frac{|A_j \cap A_k|}{|A_k|} \right], \tag{3}$$

which measures the ambiguity in terms of the amount of discrepancy among the evidential claims $m_j$. The second is the *nonspecificity*:

$$\mathbf{N}(\mathcal{S}) = \sum_j m_j \log_2(|A_j|), \tag{4}$$

which measures the "spread" of the evidence.

*This work is partly in fulfillment of the requirements for the Ph.D. degree from the State University of New York at Binghamton.

There are a number of special cases depending on the structure of $\mathcal{F}$. First, when $\forall j, |A_j| = 1$, then $\mathcal{S}$ is *specific*. We have $|\mathcal{S}| = |U|$, and $\mathrm{Bel}(A_j) = \mathrm{Pl}(A_j) = \mathrm{Pr}(A_j)$ is a probability measure with distribution $\vec{\mathrm{Pl}} = \vec{p} = \langle p_i \rangle = \langle \mathrm{Pr}(\{x_i\}) \rangle = \langle m_i \rangle$ and normalization $\sum_i p_i = 1$. The information measures are:

$$\mathbf{D}(\mathcal{S}) = \mathbf{H}(\mathcal{S}) = \mathbf{H}(\vec{p}\,) = -\sum_i p_i \log_2(p_i) \tag{5}$$

$$\mathbf{N}(\mathcal{S}) = 0 \tag{6}$$

where $\mathbf{H}$ is the stochastic entropy.

$\mathcal{S}$ is *consonant* ($\mathcal{F}$ is a *nest*) when (without loss of generality for ordering, and letting $A_0 = \emptyset$) $A_{j-1} \subset A_j$. It then follows that $|\mathcal{S}| = |U|$, and $\mathrm{Pl}(A_j) = \Pi(A_j)$ is a possibility measure. Denoting $A_i = \{x_1, x_2, \ldots, x_i\}$, and assuming that $\mathcal{S}$ is complete (i.e. $\forall x_i \in U, \exists A_i$), then the possibility distribution is $\vec{\mathrm{Pl}} = \vec{\pi} = \langle \pi_i \rangle = \langle \Pi(\{x_i\}) \rangle = \left\langle \sum_{k=i}^{|\mathcal{S}|} m_k \right\rangle$ with normalization $\bigvee_i \pi_i = 1$. For information measures, letting $\pi_{|\vec{\pi}|+1} = 0$, we have [7]:

$$\mathbf{D}(\mathcal{S}) < .892 \tag{7}$$

$$\mathbf{N}(\mathcal{S}) = \sum_j m_j \log_2(j) \tag{8}$$

$$= \mathbf{N}(\vec{\pi}) = \sum_{i=2}^{|\vec{\pi}|} \pi_i \log_2\left(\frac{i}{i-1}\right) = \sum_i (\pi_i - \pi_{i+1}) \log_2(i). \tag{9}$$

## Set-Based Statistics

The fundamental issue for constructing an empirical semantics for possibility lies in the nature of the evidence gathering. Traditionally, observations are made of the occurrence of one or another outcome $x$ out of a set of possible outcomes $U$. Denoting the count of the number of outcomes for $x_i$ as $c_i$, then for a given total count of $M$ we can arrive at a frequency distribution $f\colon U \mapsto [0,1], f(x_i) = f_i = \frac{c_i}{M}$. $\vec{f} = \langle f_i \rangle$ is a probability distribution with $\sum_i f_i = 1$ and additive measure $F(A) = \sum_{x_i \in A} f_i$.

There are a number of suggestions for conversion formulas $\vec{p} \Rightarrow \vec{\pi}$ [11, 4]. However, there can be no doubt that $\vec{f}$ and $F$ are in fact a natural probability distribution and measure with zero nonspecificity and generally positive entropy. Thus, while there may be a *good* conversion of $\vec{f} \Rightarrow \vec{\pi}$, the representation $\vec{\pi}$ is never *appropriate* for the given evidence $\vec{f}$.

Instead, what is required is to collect statistics on outcomes not in $U$, but in $2^U$. Then an observation of a subset $A \subset U$ indicates an event somewhere in $A$. Thus whenever $|A| > 1$, the observation is somewhat non-specific. We note that while researchers strive to achieve specific observations, and are frequently successful, nevertheless subset observations are in fact quite normal. In particular, subset observations result whenever the sensitivity of an instrument results in the recording of a range, or error-bars attached to a point measurement.

Denoting the count of a subset $A$ by $c_A$, then we have frequencies on subsets $f^U\colon 2^U \mapsto [0,1]$, with $f^U(A_j) = f^U_j = \frac{c_{A_j}}{M}$. $\vec{f^U}$ is a natural evidence measure generating an empirically derived random set denoted as $\mathcal{S}^E$ with focal set $\mathcal{F}^E$. When $\mathcal{S}^E$ is specific, $\vec{f^U} = \vec{f}$. When $\mathcal{S}^E$ is consonant and appropriately ordered, then $|\vec{f^U}| = |U|$, and $\vec{\pi} = \langle \mathrm{Pl}(\{x_i\}) \rangle = \left\langle \sum_{k=i}^{M} f^U_k \right\rangle$ is a possibility distribution with normalization $\bigvee_i \pi_i = 1$. In general, we have $\mathrm{Pl}_{i_1} \vee \mathrm{Pl}_{i_2} \le \mathrm{Pl}(\{x_{i_1}, x_{i_2}\}) \le \mathrm{Pl}_{i_1} + \mathrm{Pl}_{i_2}$ and $\bigvee_i \mathrm{Pl}_i \le 1 \le \sum_i \mathrm{Pl}_i$. Wang [18] has described $\vec{\mathrm{Pl}}$ as the membership grade for a fuzzy set, and set-statistics have been considered by Dubois and Prade [5, 6].

## Maximum Uncertainty Methods

Mathematical theories for the derivation of probability distributions on the basis of observed evidence have been well justified through the use of the Maximum Entropy Principle (MEP), which has been applied fruitfully to a wide variety of problems [9, 17]. Given a stochastic problem, evidence is gathered through observation, either particularly (a certain record of observations) or statistically (the observation of a certain statistical value, usually a mean). Then, using this data as a constraint, calculations are made to determine the consistent probability measure with maximum entropy, and that distribution is hypothesized as an estimate of the underlying stochastic process.

The MEP is a special case of the more general Maximum Uncertainty Principle (MUP) [10] which holds in any mathematical information theory. It states that when a problem solution is underdetermined, the possible solution with the highest uncertainty should be chosen. This generalization of Laplace's Principle of Insufficient Reason assures that the most conservative choice will be made, utilizing all available information, but no more.

So when a probability distribution is desired, then the MEP should be used. But when a possibility distribution is desired, then the MUP requires that the nonspecificity $\mathbf{N}$ should be maximized, since when $\mathcal{S}$ is consonant and $|\mathcal{S}|$ increases, $\mathbf{D}(\mathcal{S})$ tends to a small constant. We note that, in general, $\max_{\vec{\pi}} \mathbf{N}(\vec{\pi}) = \vec{1}$ with $\mathcal{S} = \{\langle U, 1 \rangle\}$.

Therefore, our method will be the following. Given an observed random set $\mathcal{S}^E$ with plausibility distribution $\vec{\mathrm{Pl}}$, we want to derive a consonant random set $\mathcal{S}'$ with focal set $\mathcal{F}'$, possibility distribution $\vec{\pi}$ and measure $\Pi$. $\mathcal{S}'$ should be "consistent" with the data $\mathcal{S}^E$, in the sense that an evidential claim $\langle A, m \rangle \in S^E$ can only be moved to an $A' \in \mathcal{F}'$ which completely accounts for that evidence, that is $A \subset A'$, and so $\mathbf{N}(\mathcal{S}') \geq \mathbf{N}(\mathcal{S}^E)$. Otherwise, when there are multiple possible solutions, the $\mathcal{S}'$ with maximal nonspecificity should be chosen. We note that this is similar to the method described in [6], where random set cardinality $|S| = \sum_j m_j |A_j|$ is used in place of $\mathbf{N}$.

## Examples

We note some simple ad-hoc examples below, with $|U| = 3$.

1. Let $\mathcal{S}^E = \{\langle \{x_1\}, a \rangle, \langle \{x_1, x_2\}, b \rangle, \langle \{x_1, x_3\}, c \rangle\}$. $\mathcal{S}^E$ is "consistent" in the sense of [5], in that $\exists \{x\} \in \mathcal{F}^E, \forall A_i, \{x\} \subset A_i$ (for us $x = x_1$). This guarantees that $\vec{\mathrm{Pl}} = \langle 1, b, c \rangle$ is already a normalized possibility distribution. Thus, assuming $b > c$, then we have $\vec{\pi} = \vec{\mathrm{Pl}} = \langle 1, b, c \rangle$, with $\mathcal{S}' = \{\langle \{x_1\}, a + c \rangle, \langle \{x_1, x_2\}, b - c \rangle, \langle \{x_1, x_2, x_3\}, c \rangle\}$. To produce a consonant $\mathcal{S}'$, either $\{x_1, x_2\}$ or $\{x_1, x_3\}$ had to be eliminated. Since $c < b$, the effect is to eliminate the lesser claim for $\{x_1, x_3\}$, by increasing the evidence in the singleton $\{x_1\}$ by $c$, and decreasing the claim of $\{x_1, x_2\}$ by the same amount. $\mathbf{N}(\mathcal{S}^E) = b + c$ increases to $\mathbf{N}(\mathcal{S}') = b + (\log_2(3) - 1)c \sim b + 1.58c$.

2. Let $\mathcal{S}^E = \{\langle \{x_1\}, a \rangle, \langle \{x_3\}, b \rangle, \langle \{x_1, x_2\}, c \rangle\}$. $\mathcal{S}^E$ is no longer consistent, so we have a subnormal $\vec{\mathrm{Pl}} = \langle a + c, c, b \rangle$. Since $\{\{x_1\}, \{x_1, x_2\}\}$ is a nest, from which $\{x_3\}$ is disjoint, then in order to construct a full nest for $\mathcal{S}'$, the claim for $\{x_3\}$ must be displaced to $U$ itself, resulting in $\mathcal{S}' = \{\langle \{x_1\}, a \rangle, \langle \{x_1, x_2\}, c \rangle, \langle U, b \rangle\}$, with $\vec{\pi} = \langle 1, b + c, b \rangle$. The effect is to add $b$ to both $\mathrm{Pl}_1$ and $\mathrm{Pl}_2$, the plausibilities of the nest in $\mathcal{S}^E$. $\mathbf{N}(\mathcal{S}^E) = c$ increases to $\mathbf{N}(\mathcal{S}') = \log_2(3)b + c$.

3. Let $\mathcal{S}^E = \{\langle \{x_1\}, a \rangle, \langle \{x_2\}, b \rangle, \langle \{x_3\}, c \rangle, \langle \{x_1, x_2\}, d \rangle\}, \vec{\mathrm{Pl}} = \langle a + d, b + d, c \rangle$. As above, $\{x_3\}$ is the disjoint set, so we move the evidence $c$ to $U$. But there are two possible nests consistent with the remaining sets, yielding two possible solutions: $\mathcal{S}'_1 = \{\langle \{x_1\}, a \rangle, \langle \{x_1, x_2\}, b + d \rangle, \langle U, c \rangle\}, \vec{\pi}_1 = \langle 1, b + c + d, c \rangle, \mathbf{N}(\mathcal{S}'_1) = \log_2(3)c + b + d$; and $\mathcal{S}'_2 = \{\langle \{x_2\}, b \rangle, \langle \{x_1, x_2\}, a + d \rangle, \langle U, c \rangle\}, \vec{\pi}_2 = \langle 1, a + c + d, c \rangle, \mathbf{N}(\mathcal{S}'_2) = \log_2(3)c + a + d$. According to maximum nonspecificity, we should clearly choose $\mathcal{S}'_1$ if $b > a$, and $\mathcal{S}'_2$ if $a > b$.

4. Assume $\mathcal{S}^E$ is specific, with a frequency distribution $\vec{f} = \langle a, b, c \rangle$. In general, there are $|U|!$ possible nests in $2^U$, 6 in our case. Each such nest yields a different nonspecificity. For example, construct $\mathcal{S}'_1$ for the nest $\{\{x_1\}, \{x_1, x_2\}, U\}$. $a$ stays with $\{x_1\}$, $b$ moves to $\{x_1, x_2\}$, and $c$ moves to $U$, yielding $\mathbf{N}(\mathcal{S}'_1) = \log_2(3)c + b$. Assuming (without loss of generality) that $a < b < c$, then $\mathbf{N}(\mathcal{S}'_1)$ is maximal for all nests, so by maximum nonspecificity it is chosen with $\vec{\pi} = \langle 1 = a + b + c, b + c, c \rangle$. We note that this recovers a known probability-possibility conversion formula [4, 13], where a fuzzy membership grade is derived as a discrete cumulative probability distribution function:

$$\mu(x) = \sum_{x_i \leq x} p_i \tag{10}$$

   except that the ordering of the $p_i$ is not determined by the structure of the space $U$, but rather by the probability values.

## Conclusion

We have examined the concept of an empirical measure of possibility derived from set-based statistics and the Principle of Maximum Uncertainty. Areas for future research include:

- Movement beyond these ad-hoc examples towards a more complete formalism of the method outlined above to general dimensionality $|U|$, and including a valid measure of "consistency" between $\mathcal{S}^E$ and $\mathcal{S}'$.

- Nonspecificity maximization relative to a possibilistic "sample mean" (analogous to the possibilistic expected value [3]) measured on a set-based data set $\mathcal{S}^E$.

- Use of the MUP to derive other algebraic structures of $\mathcal{S}'$ which yield plausibility distributions with triangular co-norm operators other than $+$ and $\vee$ [2].

# References


[1] Brillouin, Leon: (1964) *Scientific Uncertainty and Information*, Academic Press, New York

[2] Dubois, D, and Prade, H: (1982) "On Several Representations of an Uncertain Body of Evidence", *Fuzzy Information & Decision Processes*, pp. 167-181

[3] Dubois, D, and Prade, Henri: (1985) "Evidence Measures Based on Fuzzy Information", *Automatica*, v. **21**:5, pp. 547-562

[4] Dubois, D, and Prade, Henri: (1986) "Fuzzy Sets and Statistical Data", *European J. of Operational Research*, v. **25**, pp. 345-356

[5] Dubois, D, and Prade, H: (1989) "Fuzzy Sets, Probability, and Measurement", *European J. of Operational Research*, v. **40**:2, pp. 135-154

[6] Dubois, D. and Prade, H: (1990) "Consonant Approximations of Belief Functions", *Int. J. Approximate Reasoning*, v. **4**, pp. 419-449

[7] Geer, James, and Klir, George: (1991) "Discord in Possibility Theory", *Int. J. Gen. Sys.*, v. **19**

[8] Giering, Edward W, and Kandel, Abraham: (1983) "Application of Fuzzy Set Theory to the Modelling of Competition in Ecological Systems", *Fuzzy Sets and Systems*, v. **9**, pp. 103-127

[9] Jaynes, ET: (1957) "Information Theory and Statistical Mechanics", *Physical Review*, v. **106,108**, pp. 620-630

[10] Klir, George: (1990) "Uncertainty Principles in Reasoning", in: *Proc. 1990 NAFIPS*, pp. 212-214

[11] Klir, George: (1990) "Principle of Uncertainty & Information Invariance", *Int. J. Gen. Sys.*, v. **17**:2-3, pp. 249-275

[12] Klir, George, and Ramer, Arthur: (1991) "Uncertainty in Dempster-Shafer Theory: A Critical Reexamination", *Int. J. Gen. Sys.*, v. **18**

[13] MacVicar-Whalen, PJ: (1979) "Fuzzy Logic: An Alternative Approach", in: *Proc. 9th IEEE Int. Symp. on Multi-Valued Logic*, pp. 152-159, Bath

[14] Nguyen, HT: (1978) "Random Sets and Belief Functions", *J. Math. Anal. & Appl.*, v. **65**, pp. 531-542

[15] Shafer, Glen: (1976) *A Mathematical Theory of Evidence*, Princeton U., Princeton

[16] Shannon, CE, and Weaver, W: (1964) *A Mathematical Theory of Communication*, U. Illinois, Urbana

[17] Skilling, John, ed.: (1989) *Maximum-Entropy and Bayesian Methods*, Kluwer

[18] Wang, PZ: (1983) "From the Fuzzy Statistics to the Falling Random Subsets", *Advances in Fuzzy Sets, Possibility Theory, & Application*, ed. PP Wang, pp. 81-96, Plenum, New York

[19] Zadeh, Lofti A: (1978) "Fuzzy Sets as the Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, v. **1**, pp. 3-28