[12] Orlovski, S.A. (1991). Calculus of properties and fuzzy sets. *Proceedings Fourth IFSA*, Vol. Mathematics, 153-156.

[13] Schay, G. (1968). An algebra of conditional events. *J. Math. Anal. Appl.* (24), 334-344.

[14] Vitushkin, A.G. (1977). *On Representation of functions by means of superposition and related topics.* L'Enseign. Math. II, Series 23, 255-320.

[15] Wang, L.X. (1992). Fuzzy systems are universal approximators. *Proceedings First IEEE-Fuzzy Systems*, San Diego, CA, 1163-1170.

[16] Zadeh, L.A. (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.* (23), 421-427.

Joslyn, Cliff A: (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

Empirical Possibility and Minimal Information Distortion

Cliff Joslyn * †

Abstract

In order to successfully apply possibility theory to the study of physical systems, there is a need for methods of taking measurements on them which yield data governed by possibility theory. Set-based statistics are used to generate empirically derived random sets. Normal possibility distributions are available for consistent random sets, and a set of focused consistent transformations is available for inconsistent random sets. The Principle of Uncertainty Invariance is modified to provide a method which selects that consistent transformation with Minimal Information Distortion from the measured random set.

Possibilistic Models and Measurement

A **model** of a system results from the establishment of a homomorphic relation between the modeling system and the object system which it models [24]. Models may be deterministic, or may include representations of uncertainty. Probabilistic models are a staple of classical information theory [22], while possibilistic models[1] have been developed [14].

The homomorphic relation of a model involves two **measurement** operations taken on the object that yield data. These measurements correspond to **initialization** of the modeling system based on the state of the object, and then **corroboration** of the validity of the model's predictions against the future states of the object. In the context of models with uncertainty, the measured data must conform to the formalism of the uncertainty method being used.

Measurement is a fundamentally *semantic* relation: measurement procedures are not necessarily determined from the mathematical methods used, but rest on further pragmatic considerations. While the mathematical syntax of possibility theory is well established [4], its semantics are still far from such a state, leaving insufficient methods for measurement in possibilistic models:

*Graduate Fellow, Systems Science, SUNY–Binghamton, 327 Spring St. # 2, Portland ME, 04102, USA, (207) 774-0029, `cjoslyn@bingsuns.cc.binghamton.edu`, `joslyn@kong.gsfc.nasa.gov`.

†Supported under NASA Grant # NGT 50757.

[1] In this paper it will be understood that the terms "possibility value", "possibility distribution", "possibility", "possibilistic", etc. are synonymous with "membership grade", "membership function", "fuzzy", etc. as appropriate.

R. Lowen and M. Roubens (eds.), Fuzzy Logic, 143–152.

> Membership has a clear-cut formal definition. However, explicit requirements for its empirical/experimental measurement are still missing. Under these circumstances it is not surprising that apart from first steps ... genuine measurement structures have not yet been developed. [28, p. 344]

Traditional uses of possibility theory are based on the assumption that while probabilistic randomness can be measured and applied in the physical world, possibilistic nonspecificity is a result of human psychology, perception, and description [27]. This presumption has many consequences. First, the semantics of fuzziness and possibility have been overwhelmingly dominated by "linguistic variables" and similar concepts which attempt to model the subjectivity of people. Similarly, there is a preponderance of measurement methods which derive possibility values from the opinions of people (for example, "experts" in some field) [21, 28]. Lacking even this level of method, *ad hoc* possibility distributions are posited without formal justification (e.g. [9]). Finally, sometimes the issue of possibilistic measurement is avoided altogether by taking stochastic frequency data and then converting it to a possibilistic form [3, 16].

Second, while traditional (probabilistic) information theory was developed in close relation to the physics of many-body problems and thermodynamic systems [1], applications of possibilistic models are overwhelmingly in the areas of "knowledge" or "informational engineering" such as knowledge-based control systems, approximate reasoning, and decision support. Even in those attempts to apply possibilistic methods to the theory or modeling of natural, physical systems, data is predominantly collected on the basis of opinion (e.g. [9]).[2]

We are concerned with developing possibilistic models of physical systems. Thus we require measurement methods which use empirically derived data, and a semantics of possibility which is not wedded to human psychology. This paper discusses a method by which an empirically derived random set can generate a possibility distribution (see also [11, 13]). In another paper [12] we discuss measurement techniques that can generate such random sets.

Mathematical Preliminaries

We begin with the standard Dempster-Shafer evidence and possibility theories [4, 18]. Given a finite universe $\Omega = \{\omega_i\}, 1 \le i \le n$, we will call a set function $m: 2^\Omega \mapsto [0,1]$ an **evidence function**[3] 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 $1 \le j \le |\mathcal{S}| \le 2^n - 1$. We also denote the **focal set** $\mathcal{F} = \{A_j : m_j > 0\}$ and the **core** $\mathrm{C}(\mathcal{S}) = \bigcap_{\mathcal{F}} A_j$. $\mathcal{S}$ is **consistent** when $\mathrm{C}(\mathcal{S}) \ne \emptyset$. The dual **belief** and **plausibility** measures on $\forall A \subseteq \Omega$ are $\mathrm{Bel}(A) = \sum_{A_j \subseteq A} m_j$ and $\mathrm{Pl}(A) = \sum_{A_j \cap A \ne \emptyset} m_j$. We denote the **plausibility assignment** of $\mathcal{S}$ as $\vec{\mathrm{Pl}} = \langle \mathrm{Pl}(\{\omega_i\}) \rangle = \langle \mathrm{Pl}_i \rangle$.

[2] Although there are some exceptions [23].

[3] Frequently called a "basic assignment" or "basic probability assignment".

Klir and Ramer [20] identify two complementary uncertainty measures on random sets. The first is the **discord**,

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

which measures the ambiguity of $\mathcal{S}$ in terms of the amount of discrepancy among the evidential claims m_j.[4] The second is the **nonspecificity**,

$$\mathrm{N}(\mathcal{S}) = \sum_{j=1}^{|\mathcal{S}|} m_j \log_2(|A_j|),$$

which measures the "spread" of the evidence. The **total uncertainty** of a random set is then given by $\mathrm{T}(\mathcal{S}) = \mathrm{D}(\mathcal{S}) + \mathrm{N}(\mathcal{S})$. We have that $1 \le \mathrm{D}(\mathcal{S}), \mathrm{N}(\mathcal{S}), \mathrm{T}(\mathcal{S}) \le \log_2(n)$.

When $\forall A_j \in \mathcal{F}, |A_j| = 1$, then $\mathcal{S}$ is **specific**, and $\Pr(A_j) = \mathrm{Bel}(A_j) = \mathrm{Pl}(A_j)$ is an additive **probability measure**. Pr generates a **probability distribution** $\vec{p} = \vec{\mathrm{Pl}} = \langle p_i \rangle$ with additive normalization $\sum_i p_i = 1$. The information measures then are $\mathrm{D}(\mathcal{S}) = \mathrm{H}(\mathcal{S}) = -\sum_i p_i \log_2(p_i)$, where H is the stochastic entropy, and $\mathrm{N}(\mathcal{S}) = 0$.

$\mathcal{S}$ is **consonant** ($\mathcal{F}$ is a **nest**) when (without loss of generality for ordering, and letting $A_0 = \emptyset$) $A_{j-1} \subset A_j$. Now $\Pi(A_j) = \mathrm{Pl}_j$ is called a **possibility measure**. As Pr is additive, so Π is "maximal" in the sense that $\Pi\left(\bigcup_j A_j\right) = \bigvee_j \Pi(A_j)$, 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 $\vec{\pi} = \vec{\mathrm{Pl}} = \langle \pi_i \rangle$ is a **possibility distribution** with maximal normalization $\bigvee_i \pi_i = 1$. For information measures, letting $\pi_{n+1} = 0$, we have [7]:

$$\begin{aligned}
\mathrm{D}(\mathcal{S}) &= -\sum_{i=1}^{n-1} (\pi_i - \pi_{i+1}) \log_2 \left[1 - i \sum_{k=i+1}^{n} \frac{\pi_k}{k(k-1)} \right] \\
\mathrm{N}(\mathcal{S}) &= \sum_{i=2}^{n} \pi_i \log_2 \left[\frac{i}{i-1} \right] = \sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2(i).
\end{aligned}$$

It has been established [7] that in this case $\mathrm{D}(\mathcal{S})$ is bounded from above with $\lim_{|\mathcal{S}| \longrightarrow \infty} < 0.892$. Hence, possibility measures are almost discord free; their discord may often be neglected, especially when $|\mathcal{S}|$ is large.

Traditional Measurement Procedures

As discussed above, data gathering in possibilistic modeling has been dominated by two kinds of methods.

[4] Further refinements of the discord measure are currently being developed [19].

Opinion-Based Methods

There are various means by which possibility distributions are derived from the opinions of people. Sometimes a certain distribution is simply asserted based on the opinion of the researcher and the theoretical, methodological, or other *ad hoc* considerations which they bring to bear on the problem, e.g. [9]. In other cases people who have expert knowledge of the modeled system are submitted to sophisticated polling techniques to provide their opinions of the possibility values, e.g. [28, pp. 344-349]. Another common technique is called "fuzzification", in which measured crisp data is compared against a set of possibility distributions determined from some opinion method, and then aggregated to give an overall distribution of the measured data, e.g. [10].

No doubt there are situations in which such methods are either necessary or completely sufficient. For example, these methods are natural and useful when people control and intervene in system operation, and so psychological disposition is a serious factor. In other circumstances, there is a good theory of the system being modeled and little or no access to physical measurement. But these methods are unsatisfactory at best for the modeling of physical systems or other systems in which individuals do not provide direct input. Where possible, data should be derived from physical measurements in a manner which directly captures the possibilistic nature of that data.

Converted Frequencies

In stochastic models, observations are made of the occurrence of one or another outcome ω_i. Denoting that count of these occurrences as c_i, then for a given total count of M, we can arrive at a frequency distribution $f: \Omega \mapsto [0,1], f(\omega_i) = f_i = c_i/M$. Denoted as a vector, $\vec{f} = \langle f_i \rangle$ is a natural probability distribution with normalization $\sum_i f_i = 1$ and additive measure $F: 2^\Omega \mapsto [0,1]$, given by the formula $\forall A \subset \Omega, F(A) = \sum_{\omega_i \in A} f_i$.

A variety of methods are available which convert an observed frequency distribution $\vec{f}$ to a possibility distribution $\vec{\pi}$ [3, 16]. But there can be no doubt that $\vec{f}$ is in fact a natural *probability* distribution. There may be a *good* conversion $\vec{f} \Rightarrow \vec{\pi}$, and surely such a transformation must be used when *only* frequency data are available. But the representation $\vec{\pi}$ is never ultimately *appropriate* for the data gathered by a frequency distribution $\vec{f}$. It is preferable to obtain data in a form more directly similar to the ultimate possibilistic representation.

Set-Based Statistics

In a possibilistic model, data are governed by possibility distributions and possibilistic reasoning methods. While the semantics of probabilistic reasoning is based on the notions of likelihood, chance, tendency, propensity, and frequency, the semantics of possibilistic reasoning derives from notions such as similarity, compatibility, capacity, intensity, and "degree of ease" [25].

Therefore this method begins with the use of **set-based statistics** [6]. Instead of counting outcomes of the $\omega \in \Omega$, outcomes of subsets $A \subset \Omega$ are counted. An observation of a subset $A \subset \Omega$ 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 quite normal. In particular, subset observations result whenever the sensitivity of an instrument results in either the recording of a range or error-bars attached to a point measurement. This issue and other measuring methods which generate set-based statistics are discussed in [12].

Denote the count of a subset A by c_A. Then a frequency function on subsets can be constructed as $f^\Omega: 2^\Omega \mapsto [0,1]$, with $f^\Omega(A_j) = f_j^\Omega = c_A/M$. $\vec{f^\Omega}$ is a natural evidence function generating an empirically derived random set denoted as $\mathcal{S}^E$ with focal set $\mathcal{F}^E$. When $\mathcal{S}^E$ is specific, $\vec{f^\Omega} = \vec{f}$.

Consonance, Consistency, and Inclusion

The results in this section are adapted from Dubois and Prade [5]. Given an empirically derived random set $\mathcal{S}^E$, we are now concerned with its plausibility assignment $\vec{\mathrm{Pl}}$, and with deriving an empirical possibility distribution $\vec{\pi}$ based on it. It is clear that $\mathcal{S}^E$ being specific is necessary and sufficient for $\vec{\mathrm{Pl}}$ to be stochastically normal ($\sum_i \mathrm{Pl}_i = 1$), and thus a probability distribution. But $\mathcal{S}^E$ being consonant is *not* necessary for $\vec{\mathrm{Pl}}$ to be a *possibility* distribution. We have this result:

Theorem 1 *$\mathcal{S}^E$ is consistent iff $\bigvee_i Pl_i = 1$.*

Therefore it is random set consistency, not the stronger condition of consonance, which is necessary and sufficient for $\vec{\mathrm{Pl}}$ to be a possibility distribution $\vec{\pi}$.

Each possibility distribution $\vec{\pi}$ determines a possibility measure Π according to the formula $\forall A \subset \Omega, \Pi(A) = \bigvee_{\omega_i \in A} \pi_i$, and thus also determines a consonant random set constructed from Π and denoted as $\mathcal{S}^\pi$. If $\mathcal{S}^E$ is consistent but *not* consonant, then while $\vec{\mathrm{Pl}} = \vec{\pi}$ and $\bigvee_i \vec{\mathrm{Pl}} = 1$, still $\exists A_1, A_2, \mathrm{Pl}(A_1 \cup A_2) \neq \mathrm{Pl}(A_1) \vee \mathrm{Pl}(A_2)$. Thus in this case $\mathcal{S}^E \neq \mathcal{S}^\pi$, and the original random set $\mathcal{S}^E$ cannot be constructed simply from knowledge of the distribution $\vec{\pi}$.

However, we have the following result, again from [5].

Definition 1 (Weak Random Set Inclusion) *A random set $\mathcal{S}_1$ is weakly included in $\mathcal{S}_2$, denoted $\mathcal{S}_1 \subset_w \mathcal{S}_2$, when $\forall A \subset \Omega, Pl_1(A) \leq Pl_2(A)$.*

Definition 2 (Optimal Weak Inclusion) *A random set $\mathcal{S}_1$ is optimally weakly included in $\mathcal{S}_2$, denoted $\mathcal{S}_1 \subset_w^* \mathcal{S}_2$ when $\mathcal{S}_1 \subset_w \mathcal{S}_2$ and $\mathcal{S}_1$ is the maximal such random set with respect to the partial ordering $\subset_w$.*

Theorem 2 *If $\mathcal{S}^E$ is consistent, then $\mathcal{S}^\pi \subset_w^* \mathcal{S}^E$.*

Thus, for a consistent, non-consonant $\mathcal{S}^E$, we know that $\vec{\mathrm{Pl}} = \vec{\pi}$ is a possibility distribution, and that the reconstructed random set $\mathcal{S}^\pi$ is an optimal approximation to $\mathcal{S}^E$ according to this measure. Therefore we accept consistency as a sufficient criteria to generate a possibility distribution $\vec{\pi}$ from a random set $\mathcal{S}^E$.

Consistent Transformations

In a consistent random set, all the evidential claims are in partial agreement, since they all include the core. If $\mathcal{F}$ is a nest, then $\mathbf{C}(\mathcal{S}) = A_1 \in \mathcal{F}$. Therefore a consistent random set is in some sense a "partial" nest, and it is appropriate to consider possibility distributions which approximate $\mathcal{S}^E$.

But when $\mathcal{S}^E$ is not even consistent, then it is less clear what a good possibilistic approximation to it might be. However, $\mathcal{S}^E$ will need to be modified from its given form, and in a way which distorts the original structure as little as possible. We begin with the following definitions:

Definition 3 (Consistent Transformation) *A consistent transformation of a random set $\mathcal{S}$, denoted $\mathcal{S} \mapsto \hat{\mathcal{S}}$ with focal set $\hat{\mathcal{F}}$ and evidence function $\hat{m}$, moves some evidential claims $\langle A, m \rangle \in \mathcal{S}$ to $\hat{A} \in \hat{\mathcal{F}}$ such that $A \subset \hat{A}$.*

Since $A \subset \hat{A}$, all the evidence of the old claim is accounted for in the new claim $\hat{A}$. We also have that $\mathcal{S} \subset_w \hat{\mathcal{S}}$, and $\mathrm{N}(\mathcal{S}) \leq \mathrm{N}(\hat{\mathcal{S}})$. However, sometimes $\mathrm{D}(\mathcal{S}) \leq \mathrm{D}(\hat{\mathcal{S}})$, and sometimes $\mathrm{D}(\mathcal{S}) \geq \mathrm{D}(\hat{\mathcal{S}})$.

Definition 4 (Focused Consistent Transformations) *A consistent transformation focused on $\omega_i \in \Omega$ of a random set $\mathcal{S}$, denoted $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$ with focal set $\hat{\mathcal{F}}_i$ and evidence function $\hat{m}^i$, moves $\forall A_j \in \mathcal{F}$ the evidence m_j from A_j to $A_j \cup \{\omega_i\} \in \hat{\mathcal{F}}_i$.*

There is a family of n random sets $\hat{\mathcal{S}}_i$, one for each $\omega_i \in \Omega$. For a given A_j, if $\omega_i \notin A_j$, then $m(A_j)$ becomes zero while the evidence for A_j is added to the evidence of the "promoted" subset $A_j \cup \{\omega_i\}$; whereas if $\omega_i \in A_j$, then it is unchanged. Now since $\forall \hat{A}_j \in \hat{\mathcal{F}}_i, \omega_i \in \hat{A}_j$, therefore all the $\hat{\mathcal{S}}_i$ are consistent with normal possibility distributions, and generating consonant random sets. When $\mathbf{C}(\mathcal{S}) = \emptyset$, then the effect of each $\hat{\mathcal{S}}_i$ is to create a core $\mathbf{C}(\hat{\mathcal{S}}_i) = \{\omega_i\}$ with the evidential claims in $\mathcal{S}$ concentrated on the **focus** ω_i. What is required is a method to choose the "correct" focus for a given $\mathcal{S}$.

We have the following results:

Theorem 3

$$\hat{m}^i(A_j) = \begin{cases} m_j + m(A_j - \{\omega_i\}), & \omega_i \in A_j \\ 0, & \omega_i \notin A_j \end{cases}$$

Proof: Let ω_i and A_j be fixed. We can consider the losses, gains, and retentions of the evidence for A_j under the transformation $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$. The only losses will occur if $\omega_i \notin A_j$, in which case m_j is lost. If $\omega_i \in A_j$ then m_j is retained. Finally, A_j will receive gains from any A_k such that $A_k = A_j \cup \{\omega_i\}$. This is only true for $A_k = A_j$ (considered in the case of retention) or $A_k = A_j - \{\omega_i\}$. **Case 1:** Let $\omega_i \notin A_j$. Then we have the loss of m_j, no retention, and since $A_j - \{\omega_i\} = A_j$, no gains. Therefore $\hat{m}^i(A_j) = 0$. **Case 2:** Let $\omega_i \in A_j$. Then we have no losses, m_j is retained, and we gain $m(A_j - \{\omega_i\})$. Therefore $\hat{m}^i(A_j) = m_j + m(A_j - \{\omega_i\})$. ∎

Theorem 4 *$\mathcal{S} \mapsto \hat{\mathcal{S}}_i$ induces the transformation:*

$$\vec{Pl} = \langle Pl_1, Pl_2, \ldots, Pl_i, \ldots, Pl_n \rangle \mapsto \vec{\pi} = \langle Pl_1, Pl_2, \ldots, 1, \ldots, Pl_n \rangle$$

Proof: See [13].

Principles of Reasoning with Uncertainty

Klir has advanced three principles for reasoning with uncertain systems [17]. These are the principles of Minimum Uncertainty, Maximum Uncertainty, and Uncertainty Invariance.

Principle of Minimum Uncertainty: This is an arbitration principle to be used in simplification or approximation problems. It says that that solution with minimal uncertainty should be chosen so as to loose the least possible amount of information.

Principle of Maximum Uncertainty: This principle is used in the context of inductive or ampliative reasoning, when it is necessary to extrapolate beyond available information. It says that that solution with maximal uncertainty should be chosen, so that it is maximally noncommittal with regard to missing information.

Principle of Uncertainty Invariance: This principle is used when translating a problem from one formalism to another, for example probability to possibility, and requires that the quantity of uncertainty as measured in each formalism be preserved under the transformation.

In the context of probabilistic systems, the Principle of Maximum Uncertainty has had wide application as the Principle of Maximum Entropy [26]; and both the maximum and minimum entropy principles have been developed extensively by Christensen [2].

Minimal Information Distortion

In the present situation, we are interested in transforming evidence represented in the random set $\mathcal{S}^E$ to a consistent random set denoted $\mathcal{S}^*$, that is, assuming that $\mathcal{S}^E$ is not itself consistent. In this context, then, we want to apply the Principle of Uncertainty Invariance to derive $\mathcal{S}^*$ with uncertainty equal to that of $\mathcal{S}^E$. Therefore we can state the following manifestation of the Principle of Uncertainty Invariance:

Principle 1 *Given an empirically derived random set $\mathcal{S}^E$, let $\mathcal{S}^*$ be that focused, consistent transformation $\hat{\mathcal{S}}_i$ such that $\mathcal{T}(\mathcal{S}^E) = \mathcal{T}(\hat{\mathcal{S}}_i)$.*

However, Principle 1 cannot be used in this form. As the Principle of Uncertainty Invariance was originally introduced [15], one side of the transformation was considered to be completely constrained, while the other was constrained only by the measure of uncertainty. For example, for a given, fixed probability distribution, the researcher is free to select any possibility distribution with equal uncertainty. Later results [8] require specific transformation methods because of their desirable properties, but still the transformed distribution could range over a continuous parameter $\alpha \in (0, 1)$, and it was shown that $\exists \alpha \in (0, 1)$ such that uncertainty invariance could be satisfied.

But in the present context, the set of focused transformations $\{\hat{\mathcal{S}}_i\}$ provide only a *finite* set of candidates from which $\mathcal{S}^*$ must be selected. If $\mathcal{S}^E$ is already consistent then of course $\mathcal{S}^* = \mathcal{S}^E$ and so $\mathrm{T}(\mathcal{S}^*) = \mathcal{S}^E$. But in general it may very well be the case that $\nexists \hat{\mathcal{S}}_i, \mathrm{T}(\hat{\mathcal{S}}_i) = \mathrm{T}(\mathcal{S}^E)$. We know that $\forall i, \mathrm{N}(\mathcal{S}^E) < \mathrm{N}(\hat{\mathcal{S}}_i)$, but such a relation does not necessarily hold for D, and therefore also not for T. In general, there will be a tradeoff when $\mathcal{S}^E$ is transformed to $\hat{\mathcal{S}}_i$, with the discord of $\mathcal{S}^E$ being transformed into the nonspecificity of the $\hat{\mathcal{S}}_i$. But the conditions under which $\mathrm{T}(\hat{\mathcal{S}}_i)$ increases or decreases from $\mathrm{T}(\mathcal{S}^E)$ have yet to be investigated.

Therefore, we must adopt the following modification of Principle 1 in this finite case:

Principle 2 (Minimal Information Distortion) *Given an empirically derived random set $\mathcal{S}^E$, let $\mathcal{S}^*$ be that focused, consistent transformation $\hat{\mathcal{S}}_i$ such that $T(\mathcal{S}^E)$ is as "close" to $T(\hat{\mathcal{S}}_i)$ as possible.*

It is clear that we require a "distortion" function $\xi(\mathcal{S}_1, \mathcal{S}_2)$ with the following properties:

$$\xi(\mathcal{S}_1, \mathcal{S}_2) \geq 0$$

$$\xi(\mathcal{S}_1, \mathcal{S}_2) = 0 \leftrightarrow \mathrm{T}(\mathcal{S}_1) = \mathrm{T}(\mathcal{S}_2)$$

and then to select that $\hat{\mathcal{S}}_i$ for which $\xi(\mathcal{S}^E, \hat{\mathcal{S}}_i)$ is a minimum.

An obvious candidate is $\xi(\mathcal{S}_1, \mathcal{S}_2) = |\mathrm{T}(\mathcal{S}_1) - \mathrm{T}(\mathcal{S}_2)|$, but this might not always be satisfactory. Choice of a distortion function will depend on the methodology of the investigator. If $\mathrm{T}(\mathcal{S}^E) < \mathrm{T}(\hat{\mathcal{S}}_i)$, then "extra" information is gained through the transformation that was not included in the data. On the other hand, if $\mathrm{T}(\mathcal{S}^E) > \mathrm{T}(\hat{\mathcal{S}}_i)$, then information in the data is lost through the transformation. In general it should be considered more dangerous to add spurious information than to excise given information, but a very great loss should not be chosen over a very small gain. One can imagine a more sophisticated loss function which would smoothly provide more weight to information loss than information gain.

An Example

Let $\Omega = \{x, y, z\}, n = 3$, and assume the following inconsistent empirically derived random set with the following properties:

$$\mathcal{S}^E = \{\langle \{x\}, .1\rangle, \langle \{x, y\}, .7\rangle\}, \langle \{z\}, .2\rangle\}, \qquad \vec{\mathrm{Pl}} = \langle .8, .7, .2\rangle$$

$$\mathrm{N}(\mathcal{S}^E) = .7, \qquad \mathrm{D}(\mathcal{S}^E) = .805, \qquad \mathrm{T}(\mathcal{S}^E) = 1.505$$

There are three focused consistent transformations. We note that $\hat{\mathcal{S}}_z$ is actually consonant, and the only one for which $\mathrm{T}(\hat{\mathcal{S}}_i)$ increases (slightly). We use $\xi(\mathcal{S}^E, \hat{\mathcal{S}}_i) = |\mathrm{T}(\mathcal{S}^E) - \mathrm{T}(\hat{\mathcal{S}}_i)|$, and would thus choose $\mathcal{S}^* = \hat{\mathcal{S}}_z$.

$$\hat{\mathcal{S}}_x = \{\langle \{x\}, .1\rangle, \langle \{x, y\}, .7\rangle, \langle \{x, z\}, .2\rangle\}, \qquad \vec{\pi}_x = \langle 1, .7, .2\rangle$$

$$\mathrm{N}(\hat{\mathcal{S}}_x) = .9, \qquad \mathrm{D}(\hat{\mathcal{S}}_x) = .317, \qquad \mathrm{T}(\hat{\mathcal{S}}_x) = 1.21, \qquad \xi(\mathcal{S}^E, \hat{\mathcal{S}}_x) = .288$$

$$\hat{\mathcal{S}}_y = \{\langle \{x, y\}, .8\rangle, \langle \{x, z\}, .2\rangle\}, \qquad \vec{\pi}_y = \langle .8, 1, .2\rangle$$

$$\mathrm{N}(\hat{\mathcal{S}}_y) = 1, \qquad \mathrm{D}(\hat{\mathcal{S}}_y) = .269, \qquad \mathrm{T}(\hat{\mathcal{S}}_y) = 1.269, \qquad \xi(\mathcal{S}^E, \hat{\mathcal{S}}_y) = .236$$

$$\hat{\mathcal{S}}_z = \{\langle \{z\}, .2\rangle, \langle \{x, z\}, .1\rangle, \langle \Omega, .7\rangle\}, \qquad \vec{\pi}_z = \langle .8, .7, 1\rangle$$

$$\mathrm{N}(\hat{\mathcal{S}}_z) = 1.209, \qquad \mathrm{D}(\hat{\mathcal{S}}_z) = .320, \qquad \mathrm{T}(\hat{\mathcal{S}}_z) = 1.529, \qquad \xi(\mathcal{S}^E, \hat{\mathcal{S}}_y) = .024$$

Figure 1 shows a graphical representation of the random sets.

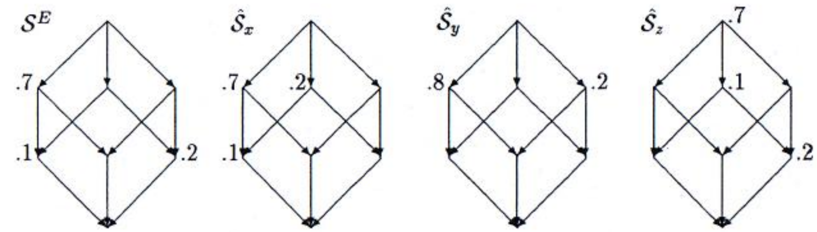

Figure 1: The power set 2^Ω is represented as a three-dimensional hypercube. The directed arcs represent subset inclusion. Numbered nodes indicate values for m, with $m = 0$ for unnumbered nodes.

References

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

[2] Christensen, Ronald: (1980) *Entropy Minimax Sourcebook*, v. **1-4**, Entropy Limited, Lincoln, MA

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

[4] Dubois, Didier and Prade, Henri: (1988) *Possibility Theory*, Plenum Press, New York

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

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

[7] Geer, James, and Klir, George: (1991) "Discord in Possibility Theory", *Int. J. General Systems*, v. **19**, pp. 119-132

[8] Geer, James, and Klir, George: (1992) "A Mathematical Analysis of Information Preserving Transformations Between Probabilistic and Possibilistic Formulations of Uncertainty", *Int. J. General Systems*, v. **20**, pp. 143-176

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

[10] Gurocak, HB and Lazaro, A de Sam: (1992) "Fuzzy Logic Approaches for Handling Imprecise Measurements in Robot Assembly", in: *Proc. 1992 FUZZ-IEEE Conference*, ed. Jim Bezdek, pp. 915-922, IEEE, San Diego

[11] Joslyn, Cliff: (1991) "Towards an Empirical Semantics of Possibility Through Maximum Uncertainty", in: *Proc. IFSA 1991*, vol. **A**, ed. R. Lowen, M. Roubens, p. 86-89, IFSA, Brussels

[12] Joslyn, Cliff: (1992) "Possibilistic Measurement and Set Statistics", submitted to *1992 NAFIPS Conference*

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

[14] Kandel, Abraham: (1986) *Fuzzy Mathematical Techniques with Applications*, Addison-Wesley

[15] Klir, George: (1989) "Probability-Possibility Conversion", *Proc. 3rd IFSA Congress*, pp. 408-411

[16] Klir, George: (1990) "A Principle of Uncertainty and Information Invariance", *Int. J. General Systems*, v. **17**, pp. 249-275

[17] Klir, George: (1991) "Measures and Principles of Uncertainty and Information: Recent Development", in: *Information Dynamics*, ed. H. Atmanspacher, pp. 1-14, Plenum Press, New York

[18] Klir, George and Folger, Tina: (1987) *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall

[19] Klir, George and Parviz, Behvad: (1992) "Note on the Measure of Discord", to appear in: *Proc. 8th Conference on Uncertainty in Artifical Intelligence*, Stanford

[20] Klir, George, and Ramer, Arthur: (1991) "Uncertainty in Dempster-Shafer Theory: A Critical Reexamination", *Int. J. General Systems*, v. **18**, pp. 155-166

[21] Norwich, AM and Turksen, IB: (1982) "Meaningfulness in Fuzzy Set Theory", in: *Fuzzy Sets and Possibility Theory*, ed. R. Yager, pp. 68-74, Pergamon, Oxford

[22] Ripley, Brian D: (1987) *Stochastic Simulation*, Wiley, New York

[23] Roberts, DW: (1989) "Analysis of Forest Succession with Fuzzy Graph Theory", *Ecological Modeling*, v. **45**, pp. 261-274

[24] Rosen, Robert: (1985) *Anticipatory Systems*, Pergamon, Oxford

[25] Ruspini, EH: (1989) "The Semantics of Vague Knowledge", *Revue Internationale de Systemique*, v. **3**, pp. 387-420

[26] Skilling, J, ed.: (1989) *Maximum-Entropy and Bayesian Methods*, Kluwer, New York

[27] Zadeh, Lofti A: (1965) "Fuzzy Sets", *Information and Control*, v. **8**, pp. 338-353

[28] Zimmerman, HJ: (1991) *Fuzzy Set Theory*, 2nd edition, Kluwer, Boston

ON CO-OPTIMAL LIFTS IN THE CATEGORY OF FUZZY NEIGHBOURHOOD SPACES

E.E. KERRE
Dept. of Mathematics
State University of Gent
Gent, Belgium

Nehad N. MORSI
Dept. of Mathematics
Military Technical College
Cairo, Egypt

We consider a subset M of a universe X as a *crisp fuzzy set*, and denote this again by M. We denote by $\underline{\alpha}$ the constant fuzzy subset of X with value $\alpha \in I$. We denote the strong α-cut of a fuzzy set $\mu \in I^X$ by $\mu^{\alpha} (\in 2^X)$. We follow Lowen's definition of the category FTS of fuzzy topological spaces (fts's) and their continuous functions. All categories appearing below will be categories of sets with structures, and all functors between them will preserve sets underlying their objects, and will be identities on morphisms. Therefore, functors will be defined through their object functions only.

The full subcategory of FTS of fuzzy neighbourhood spaces (fns's) is denoted by FNS. Wuyts [11] has shown that a fns is uniquely determined by its *topology spectrum* (= its indexed family of *level topologies*).

The α-*level functors*, $\alpha \in I_1$, $\tilde{\iota}_{\alpha} : \mathrm{FTS} \to \mathrm{TOP} : (X,\tau) \mapsto (X, \iota_{\alpha}(\tau))$ preserve optimality [3]. But neither they nor their restrictions to FNS preserve co-optimality.

In [5], it is shown that both FTS and FNS are topological categories, and that FNS is closed under the formation of optimal lifts in FTS. Also, the fuzzy neighbourhood system of an optimal lift in FNS is described there in terms of the fuzzy neighbourhood systems of the source fns's.

R. Lowen and M. Roubens (eds.), Fuzzy Logic, 153–159.