"Minimal Information Loss Possibilistic Approximations of Random Sets", in: Proc. 1992 IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE 92), ed. Jim Bezdek, pp. 1081-1088, IEEE, San Diego, https://doi.org/10.1109/FUZZY.1992.258698

Minimal Information Loss Possibilistic Approximations of Random Sets

Cliff Joslyn *†
Systems Science, SUNY–Binghamton

George Klir ‡
Systems Science, SUNY–Binghamton

Janurary 1992

Abstract

We suggest an empirical measuring procedure which yields data governed by possibility theory. Such methods are needed in order to successfully apply possibility theory to the study of physical systems. Set-based statistics are used to generate empirically derived random sets. Normal possibility distributions are available for all consistent random sets, and a set of "consistent transformations" are available for all inconsistent random sets. The Principle of Uncertainty Invariance is used in a modified form to select the consistent transformation with minimal information loss from the original random set.

1 Possibilistic Models

In a previous work [8], Joslyn considered the need for an empirically based semantics for possibility measures and distributions. Traditional methods for obtaining fuzzy set membership and possibility values may be sufficient for human-controlled, knowledge-based applications (such as control and expert systems), but an empirical approach will be needed if possibility theory is to be successfully applied to such "non-traditional" areas as the modeling of complex physical systems.

A model of a physical system, in the most general sense, requires data provided by measurements. Such measurements provide a semantic relation between the model and its object, or referent. Measured input data serve to initialize the model; while generated output data corroborate the model against further measurements.

In a stochastic model, input and output data are governed by probability theory. Measured data are converted to frequency distributions with an additive measure, and are thereafter considered as probability distributions. Monte Carlo methods are then used to generate output data according to probability distributions inherent to the model.

In a fuzzy or 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, frequency, and the like, the semantics of possibilistic reasoning derives from notions such as similarity, resemblance, elasticity, intensity, and degree of ease [15]. In this paper we suggest a measurement procedure to generate possibility distributions from non-frequency data.

*Supported under NASA Grant # NGT 50757.

†327 Spring St. # 2, Portland, ME 04102, USA; cjoslyn@bingvaxu.cc.binghamton.edu, joslyn@kong.gsfc.nasa.gov.

‡Systems Science, SUNY–Binghamton, Binghamton, NY 13901, USA.

2 Mathematical Preliminaries

First, let us overview the standard evidence theory. For a finite universe $\Omega = \{\omega_i\}, 1 \leq i \leq n$ with power set $\mathcal{P}(\Omega) = \{A \subset \Omega\}, m: \mathcal{P}(\Omega) \mapsto [0,1]$ is an evidence function on the subsets of Ω with $m(\emptyset) = 0$ and $\sum_{A \subset \Omega} m(A) = 1$. 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, $A_j \subset \Omega, m_j = m(A_j)$, and $1 \leq j \leq |\mathcal{S}| \leq 2^n - 1$. We also have the focal set $\mathcal{F} = \{A_j : m_j > 0\}$. Then the dual belief and plausibility measures on an $A \subset \Omega$ are

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

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

Following [8], we denote the "plausibility assignment" of a random set $\mathcal{S}$ as $\vec{\mathrm{Pl}} = \langle \mathrm{Pl}(\{\omega_i\}) \rangle = \langle \mathrm{Pl}_i \rangle$.

Klir and Ramer [12, 14] identify two complementary uncertainty measures on random sets. The first is the *discord*,

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

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

$$\mathrm{N}(\mathcal{S}) = \sum_{j=1}^{|\mathcal{S}|} m_j \log_2(|A_j|), \tag{4}$$

which measures the "spread" of the evidence. The *total uncertainty* of a random set is then given by

$$\mathrm{T}(\mathcal{S}) = \mathrm{D}(\mathcal{S}) + \mathrm{N}(\mathcal{S}). \tag{5}$$

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

$$\mathrm{D}(\mathcal{S}) = \mathrm{H}(\mathcal{S}) = -\sum_{i=1}^{n} p_i \log_2(p_i), \tag{6}$$

$$\mathrm{N}(\mathcal{S}) = 0, \tag{7}$$

where H is the stochastic entropy.

$\mathcal{S}$ is *consonant* ($\mathcal{F}$ is a *nest*) when (without loss of generality for ordering, and letting $A_0 = \emptyset$) $A_{j-1} \subset A_j$. As with the probabilistic case, $|\mathcal{S}| = n$, but $\mathrm{Pl}(A_j) = \Pi(A_j)$ is a *possibility measure.* Denoting $A_i = \{\omega_1, \omega_2, \ldots, \omega_i\}$, and assuming that $\mathcal{S}$ is complete (i.e. $\forall \omega_i \in \Omega, \exists A_i$), then the *possibility distribution* is $\vec{\mathrm{Pl}} = \vec{\pi} = \langle \pi_i \rangle = \langle \Pi(\{\omega_i\}) \rangle = \left\langle \sum_{k=i}^{|\mathcal{S}|} m_k \right\rangle$ with normalization $\bigvee_{i=1}^n \pi_i = 1$, where $\vee$ is maximization. For information measures, letting $\pi_{n+1} = 0$, we have [5]:

$$\mathrm{D}(\mathcal{S}) = -\sum_{j=1}^{|\mathcal{S}|-1} m_j \log_2 \left[\sum_{k=1}^{j} m_k + \sum_{k=j+1}^{|\mathcal{S}|} m_k \frac{j}{k} \right] \tag{8}$$

$$= -\sum_{i=1}^{n-1}(\pi_i - \pi_{i+1})\log_2\left[1 - i\sum_{k=i+1}^{n}\frac{\pi_k}{k(k-1)}\right] \quad (9)$$

$$\mathrm{N}(\mathcal{S}) = \sum_{j=1}^{|\mathcal{S}|} m_j \log_2(j) \quad (10)$$

$$= \sum_{i=2}^{n}\pi_i \log_2\left[\frac{i}{i-1}\right] = \sum_{i=1}^{n}(\pi_i - \pi_{i+1})\log_2(i). \quad (11)$$

It has been established [5] that the maximum values for $\mathrm{D}(\mathcal{S})$ for possibility measures is bounded from above, and the actual upper bound (for $|\mathcal{S}| \longrightarrow \infty$) is approximately 0.892. Hence, possibility measures are almost discord free; their discord may often be neglected, especially when $|\mathcal{S}|$ is large.

3 Traditional Measurement Procedures

Data-gathering in fuzzy modeling has been dominated by two kinds of methods.

3.1 Opinion-Based Methods

There are various means by which possibility or fuzzy membership values are derived from the opinions of people. Sometimes researchers assume certain distributions based on theoretical, methodological, or other *ad hoc* considerations which are outside of the model *per se*, e.g. [7]. Sometimes people who have expert knowledge of the modeled system are polled to provide their opinions of the values of the possibility of the various $\omega \in \Omega$, e.g. [13].

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, fuzzy data should be derived from physical measurement.

3.2 Converted Frequencies

In stochastic models, observations are made of the occurrence of one or another outcome ω_i. Denoting that count of these occurences as c_i, then for a given total count of M, we can arrive at a frequency distribution $f\colon \Omega \mapsto [0,1], f(\omega_i) = f_i = \frac{c_i}{M}$. Denoted as a vector, $\vec{f} = \langle f_i \rangle$ is a natural probability distribution with normalization $\sum_{i=1}^{n} f_i = 1$ and additive measure $F\colon \mathcal{P}(\Omega) \mapsto [0,1]$, 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 $\vec{\pi}$ [2, 10]. 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.

4 Set-Based Statistics

The method begins with the use of *set-based statistics* [3, 17]. 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 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. Subset observations also result whenever an observation is made from a continuous scale; and may result from observations on a discrete scale with a high resolution.

Denote the count of a subset A by c_A. Then a frequency function on subsets can be constructed as $f^\Omega : \mathcal{P}(\Omega) \mapsto [0,1]$, with $f^\Omega(A_j) = f_j^\Omega = \frac{c_A}{M}$. $\vec{f^\Omega}$ 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, $\vec{f^\Omega} = \vec{f}$. When $\mathcal{S}^E$ is consonant and appropriately ordered, then $|\vec{f^\Omega}| = n$, and $\vec{\pi} = \langle \mathrm{Pl}(\{\omega_i\}) \rangle = \left\langle \sum_{k=i}^{M} f_k^\Omega \right\rangle$ is a possibility distribution with normalization $\bigvee_{i=1}^n \pi_i = 1$.

5 Consonance, Consistency, and Inclusion

Given an empirically derived random set $\mathcal{S}^E$, we are now concerned with its plausibility assignment $\vec{\mathrm{Pl}}$, and in deriving an empirical possibility distribution $\vec{\pi}$ based on it. It is clear that $\mathcal{S}^E$ being specific is necessary and sufficient for $\vec{\mathrm{Pl}}$ being stochastically normal ($\sum_{i=1}^n \mathrm{Pl}_i = 1$), and thus a probability distribution. But $\mathcal{S}^E$ being consonant is not necessary for $\vec{\mathrm{Pl}}$ to be a possibility distribution. The following is adapted from Dubois and Prade [4].

Definition 1 (Core) *The core $C(\mathcal{S})$ of a random set $\mathcal{S}$ is $\bigcap_{\mathcal{F}} A_j$.*

Definition 2 (Random Set Consistency) *A random set is consistent when $C(\mathcal{S}) \neq \emptyset$.*

Since $\omega_i \in C(\mathcal{S}) \rightarrow \forall A_j, \{\omega_i\} \cap A_j \neq \emptyset \rightarrow \mathrm{Pl}(\{\omega_i\}) = 1$, random set consistency entails that $\vec{\mathrm{Pl}}$ is a possibility distribution. On the other hand, if $\vec{\mathrm{Pl}}$ is a possibility distribution, then $\exists \omega_i, \mathrm{Pl}(\{\omega_i\}) = 1 = \sum_{\omega_i \in A_j} m_j$. Since $\sum_{j=1}^{|\mathcal{S}|} m_j = 1$, therefore $\forall A_j \in \mathcal{F}, \omega_i \in A_j$, and $\mathcal{S}$ is consistent. Therefore it is random set consistency, not 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 and thus a consonant random set denoted $\mathcal{S}^\pi$. If $\mathcal{S}^E$ is consistent but *not* consonant, then while $\vec{\mathrm{Pl}} = \vec{\pi}$ and $\bigvee_{i=1}^n \vec{\mathrm{Pl}} = 1$, 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 [4].

Definition 3 (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 4 (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 1 *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 $\vec{\mathrm{Pl}}$ 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.

6 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 $C(\mathcal{S}) = A_1 \in \mathcal{F}$. 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 its good possibilistic approximations 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 5 (Consistent Transformation) *A consistent transformation of a random set* $\mathcal{S}$, *denoted* $\mathcal{S} \mapsto \hat{\mathcal{S}}$, *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 6 (Focused Consistent Transformations) *A consistent transformation focused on* $\omega_i \in \Omega$ *of a random set* $\mathcal{S}$, *denoted* $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$, *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 $\omega_i \in \Omega$. The effect of each $\hat{\mathcal{S}}_i$ is to "focus" the evidential claims in $\mathcal{S}$ onto the element ω_i. If $\omega_i \notin A$, then $m(A)$ becomes zero while the evidence for A is added to the evidence of the "promoted" subset $A \cup \{\omega_i\}$; whereas if $\omega_i \in A$, then it is unchanged. Now since $\forall \hat{A} \in \hat{\mathcal{F}}_i, \omega_i \in \hat{A}$, therefore all the $\hat{\mathcal{S}}_i$ are consistent with normal possibility distributions, and generating consonant random sets denoted $\hat{\mathcal{S}}_i^\pi$.

Theorem 2 $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$ *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: Let i be fixed; let $A \in \mathcal{F}, \hat{A} \in \hat{\mathcal{F}}_i$; let m, Pl and $\hat{m}, \widehat{\mathrm{Pl}}$ be the evidence functions and plausibilities of $\mathcal{S}$ and $\hat{\mathcal{S}}_i$ respectively; and let $\vec{\pi}$ be the possibility distribution of $\hat{\mathcal{S}}_i$. First, we know that $\pi_i = \widehat{\mathrm{Pl}}(\{\omega_i\}) = \sum_{\omega_i \in \hat{A}} m(\hat{A}) = \sum_{\hat{A} \in \hat{\mathcal{F}}} m(\hat{A}) = 1$. Now consider $\forall k \neq i, 1 \leq k \leq n$, and any $A_0 \in \mathcal{F}$. **Case 1:** If $\omega_i \in A_0$, then $m(A_0)$ is unchanged in the transformation. **Case 2:** Assume $\omega_i \notin A_0$. If $\omega_k \in A_0$, then $\omega_k \in A_0 \cup \{\omega_i\}$; as $m(A_0)$ is added to Pl_k, so $\hat{m}(A_0 \cup \{\omega_i\})$ is added to $\widehat{\mathrm{Pl}}_k$. Similarly, if $\omega_k \notin A_0$, then $\omega_k \notin A_0 \cup \{\omega_i\}$; as $m(A_0)$ is *not* added to Pl_k, so $\hat{m}(A_0 \cup \{\omega_i\})$ is not added to $\widehat{\mathrm{Pl}}_k$. Therefore the transformation does not change the value of $\mathrm{Pl}(\{\omega_k\})$, and $\pi_k = \widehat{\mathrm{Pl}}_k = \mathrm{Pl}_k$. ■

7 Principles of Reasoning with Uncertainty

Three principles based on relevant measures of uncertainty are fundamental to reasoning under uncertainty [11]. These are the principles of Minimum Uncertainty, Maximum Uncertainty, and Uncertainty Invariance.

Principle of Minimum Uncertainty: This is an arbitration principle, which helps us to choose solutions in simplification or approximation problems that involve uncertainty. This principle requires that we accept only those solutions, from among all otherwise acceptable solutions, whose uncertainty (pertaining to the purpose involved) is minimal. By using this principle, we accomplish a desired simplification or approximation by loosing the least possible amount of information.

Principle of Maximum Uncertainty: This principle helps us to properly extrapolate beyond the information available. The principle requires that any extrapolation be obtained by maximizing relevant uncertainty within the constraints expressing the available information. This principle guarantees that our extrapolations are maximally noncommittal with regard to missing information.

Principle of Uncertainty Invariance: This principle attempts to establish connections among representations of uncertainty in alternative mathematical theories. It requires that the amount of uncertainty be preserved when we transfer uncertainty formalized in one mathematical theory into an equivalent formalization in another theory. That is, the principle guarantees that no information is unwittingly added or eliminated solely by changing the mathematical framework in which a problem is formalized.

In the context of probabilistic systems, the Principle of Maximum Uncertainty has had wide application as the Principle of Maximum Entropy [16]; and both the maximum and minimum entropy principles have been developed extensively by Christensen [1].

8 Minimal Information Loss

In the present situation, we are interested in transforming evidence represented by the random set $\mathcal{S}^E$ to a consonant random set $\mathcal{S}^\pi$ determined by a possibility distribution $\vec{\pi}$ derived from $\mathcal{S}^E$. In this context, then, we want to apply the Principle of Uncertainty Invariance to derive a $\vec{\pi}$ with uncertainty equal to that of the original random set.

We can state the manifestation of the Principle of Uncertainty Invariance in this context as:

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

However, Principle 1 cannot be used in this form. As the Principle of Uncertainty Invariance was originally introduced [9], 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 [6] 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 possibilistic approximations $\{\hat{\mathcal{S}}_i^\pi\}$ provide only a *finite* set of candidates from which the optimal consonant random set $\mathcal{S}^\pi$ must be selected. Further, it may very well be the case that $\not\exists \hat{\mathcal{S}}_i^\pi, \mathrm{T}(\hat{\mathcal{S}}_i^\pi) = \mathrm{T}(\mathcal{S}^E)$. We know that $\forall i, \mathrm{N}(\mathcal{S}^E) < \mathrm{N}(\hat{\mathcal{S}}_i^\pi)$, 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 $\hat{\mathcal{S}}_i^\pi$, with the discord of $\mathcal{S}^E$ being transformed into the nonspecificity of the $\hat{\mathcal{S}}_i^\pi$. But the conditions under which $\mathrm{T}(\hat{\mathcal{S}}_i^\pi)$ 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 Loss) *Given an empirically derived random set $\mathcal{S}^E$, let $\mathcal{S}^\pi$ be that focused, consistently transformed possibilistic approximation $\hat{\mathcal{S}}_i^\pi$ such that $T(\mathcal{S}^E)$ is as "close" to $T(\hat{\mathcal{S}}_i^\pi)$ as possible.*

It is clear that we require an "information loss" function $L(\mathcal{S}_1, \mathcal{S}_2)$ with $L(\mathcal{S}_1, \mathcal{S}_2) = 0 \leftrightarrow T(\mathcal{S}_1) = T(\mathcal{S}_2)$, and then to select that $\hat{\mathcal{S}}_i^\pi$ for which $L(\mathcal{S}^E, \hat{\mathcal{S}}_i^\pi)$ is a minimum. An obvious candidate is $L(\mathcal{S}_1, \mathcal{S}_2) = |T(\mathcal{S}_1) - T(\mathcal{S}_2)|$, but this might not always be satisfactory.

Choice of a loss function will depend on the methodology of the investigator. If $T(\mathcal{S}^E) < T(\hat{\mathcal{S}}_i^\pi)$, then "extra" information is gained through the transformation that was not included in the data. On the other hand, if $T(\mathcal{S}^E) > T(\hat{\mathcal{S}}_i^\pi)$, 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.

References

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

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

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

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

[5] Geer, J, and Klir, G: (1991) "Discord in Possibility Theory", *Int. J. Gen. Sys.*, v. **19**, pp. 119-132

[6] Geer, J, and Klir, G: (1992) "A Mathematical Analysis of Information Preserving Transformations Between Probabilistic and Possibilistic Formulations of Uncertainty", *Int. J. Gen. Sys.*, v. **20**, in press

[7] 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

[8] Joslyn, C: (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

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

[10] Klir, G: (1990) "A Principle of Uncertainty and Information Invariance", *Int. J. Gen. Sys.*, v. **17**, pp. 249-275

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

[12] Klir, G, and Ramer, A: (1991) "Uncertainty in Dempster-Shafer Theory: A Critical Reexamination", *Int. J. Gen. Sys.*, v. **18**, pp. 155-166

[13] Murase, Y, and Usui, M et. al.: (1991) "An Application of Fuzzy Control to Waste Water Pumping Station", in: *Proc. IFSA '91*, v. **E**, ed. R. Lowen, pp. 131-134, IFSA, Brussels

[14] Ramer, A: (1991) "Inequalities in Evidence Theory Based on Concordance and Conflict", in: *Proc. IFSA '91*, v. **A**, ed. R. Lowen, M. Roubens, pp. 172-175, IFSA, Brussels

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

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

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