RANDOM FUZZY SETS

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
phone (505) 646-2105
fax (505) 646-1064
hunguyen@nmsu.edu

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
phone (915) 747-6951
fax (915) 747-5030
vladik@utep.edu

Gang Xiang
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
phone (915) 799-3621
fax (915) 747-5030
gxiang@acm.org

RANDOM FUZZY SETS

ABSTRACT

It is well known that in decision making under uncertainty, while we are guided by a general (and abstract) theory of probability and of statistical inference, each specific type of observed data requires its own analysis. Thus, while textbook techniques treat precisely observed data in multivariate analysis, there are many open research problems when data are censored (e.g., in medical or bio-statistics), missing, or partially observed (e.g., in bioinformatics). Data can be imprecise due to various reasons, e.g., due to fuzziness of linguistic data. Imprecise observed data are usually called *coarse data*. In this chapter, we consider coarse data which are both random and fuzzy.

Fuzziness is a form of imprecision often encountered in perception-based information. In order to develop statistical reference procedures based on such data, we need to model random fuzzy data as bona fide random elements, i.e., we need to place random fuzzy data completely within the rigorous theory of probability. This chapter presents the most general framework for random fuzzy data, namely the framework of random fuzzy sets. We also describe several applications of this framework.

Keywords: continuous lattices, random sets, random fuzzy sets, Choquet theorem, coarse data, perception-based information

FROM MULTIVARIATE STATISTICAL ANALYSIS TO RANDOM SETS

What is a random set? An intuitive meaning. What is a random set? Crudely speaking, a random number means that we have different numbers with different probabilities, a random vector means that we have different vectors with different probabilities; similarly, a random set means that we have different sets with different probabilities.

How can we describe this intuitive idea in precise terms? To provide such a formalization, let us recall how probabilities and random vectors are usually defined.

How probabilities are usually defined. To describe probabilities, in general, we must have a set Ω of possible situations $\omega \in \Omega$, and we must be able to describe the probability P of different properties of such situations. In mathematical terms, a *property* can be characterized by the *set* of all the situations ω which satisfy this property. Thus, we must assign to sets $A \subseteq \Omega$, the probability value $P(A)$.

According to the intuitive meaning of probability (e.g., as frequency), if we have two disjoint sets A and A', then we must have $P(A \cup A') = P(A) + P(A')$. Similarly, if we have countably many mutually disjoint sets A_i, we must have $P\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{\infty} P(A_i)$. A mapping which satisfies this property is called *σ-additive*.

It is known that even in the simplest situations, e.g., when we randomly select a number from the interval $\Omega = [0, 1]$, it is not possible to have a σ-additive function P which would be defined on *all* subsets of $[0, 1]$. Thus, we must restrict ourselves to a class $\mathcal{A}$ of subsets of Ω. Since subsets represent properties, a restriction on subsets means restriction on properties. If we allow two properties F and F', then we should also be able to consider their logical

combinations $F \,\&\, F'$, $F \vee F'$, and $\neg F$ – which in set terms correspond to union, intersection, and complement. Similarly, if we have a sequence of properties F_n, then we should also allow properties $\forall n\, F_n$ and $\exists n\, F_n$ which correspond to countable union and intersection. Thus, the desired family $\mathcal{A}$ should be closed under (countable) union, (countable) intersection, and complement. Such a family is called a *σ-algebra.*

Thus, we arrive at a standard definition of a *probability space* as a triple $(\Omega, \mathcal{A}, P)$, where Ω is a set, $\mathcal{A}$ is a σ-algebra of subsets of Ω, and $P : \mathcal{A} \to [0, 1]$ is a σ-additive function.

How random objects are usually defined. Once a probability space is fixed, a random object from the set $\mathcal{C}$ is defined as mapping $V : \Omega \to \mathcal{C}$. For example, the set of all d-dimensional vectors is the Euclidean space $\mathbb{R}^d$, so a random vector is defined as a map $V : \Omega \to \mathbb{R}^d$.

The map V must enable us to define probabilities of different properties of objects. For that, we need to fix a class of properties; we already know that such a class $\mathcal{B}$ should be a σ-algebra. For each property of objects, i.e., for each set $B \in \mathcal{B}$, it is natural to define the probability $P(B)$ as the probability that for a random situation ω, the corresponding object $V(\omega)$ belongs to the set B (i.e., satisfies the desired property). In precise terms, this means that we define this probability as $P(V^{-1}(B))$, where $V^{-1}(B)$ denotes $\{\omega : V(\omega) \in B\}$.

For this definition to be applicable, we must require that for every set B from the desired σ-algebra $\mathcal{B}$, the set $V^{-1}(B)$ must belong to $\mathcal{A}$. Such mappings are called *$\mathcal{A}$-$\mathcal{B}$-measurable.*

How random vectors are usually defined. In particular, for random vectors, it is reasonable to allow properties corresponding to all open sets $B \subseteq \mathbb{R}^d$ (like $x_1 > 0$) and properties corresponding to all closed sets B (such as $x_1 \geq 0$). So, we must consider the smallest σ-algebra which contains all closed and open sets. Sets from this smallest subalgebra are called *Borel sets*; the class of all Borel sets over $\mathbb{R}^d$ is denoted by $\mathcal{B}(\mathbb{R}^d)$. So, a random vector is usually defined as a $\mathcal{A}$-$\mathcal{B}^d$-measurable mapping $V : \Omega \to \mathbb{R}^d$.

Alternatively, we can consider the vectors themselves as events, i.e., consider a probability space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_V)$. In this reformulation, as we have mentioned, for every set B, we get $P_V(B) = P(V^{-1}(B))$, so we can say that P_V is a composition of the two mappings: $P_V = PV^{-1}$. This measure P_V is called a *probability law* of the random vector.

How random vectors are usually described. From the purely mathematical viewpoint, this is a perfect definition of a random vector. However, from the computational viewpoint, the need to describe $P_V(B)$ for all Borel sets B makes this description impractical. Good news is that due to σ-additivity, we do not need to consider all possible Borel sets B, it is sufficient to describe $P_V(B)$ only for the sets of the type $(-\infty, x_1] \times \ldots \times (-\infty, x_d]$, i.e., it is sufficient to consider only the probabilities that $\xi_1 \leq x_1 \,\&\, \ldots \,\&\, \xi_d \leq x_d$.

In the 1-dimensional case, the probability $F(x)$ that $\xi \leq x$ is called a (cumulative) distribution function. Similarly, in the general d-dimensional case, the probability $F(x_1, \ldots, x_d)$ that $\xi_1 \leq x_1$, …, and $\xi_d \leq x_d$ is called a distribution function. In these terms, the probability measures on $\mathcal{B}(\mathbb{R}^d)$ are uniquely characterized by their distribution functions. This result was first proved by Lebesgue and Stieltjes and is thus called the *Lebesgue-Stieltjes theorem.*

From random vectors to slightly more general random objects. The above definition of a random vector does not use any specific features of a multi-dimensional vector space $\mathbb{R}^d$, so it can be naturally generalized to the case when the set $\mathcal{C}$ of objects (possible outcomes) is a locally compact Hausdorff second countable topological space. Such spaces will be described as LCHS, for short.

From random vectors to random sets. As we have mentioned, in some real-life situations, the outcome is not a vector by a set of possible vectors. In this case, possible outcomes are subsets of the vector space $\mathbb{R}^d$, so the set $\mathcal{C}$ of possible outcomes can be described as the set of all such subsets – a *power set* $\mathcal{P}(\mathbb{R}^d)$.

Need to consider one-element sets. An important particular case of this general situation is the case when we know the exact vector x. In this case, the set of all possible vectors is the corresponding one-element set $\{x\}$. So, one-element sets must appear as possible sets.

Restriction to closed sets. From the practical viewpoint, it is sufficient to only consider *closed sets*.

Indeed, by definition, a closed set is a set which contains all the limits of its elements. If $x_n \in S$ and $x_n \to x$, then, by definition of a limit, this means that whatever accuracy we choose, we cannot distinguish between x and values x_n for sufficiently large n. Since $x_n \in S$ and x is undistinguishable from x_n, it makes sense to conclude that x also belongs to S – i.e., that S is indeed a closed set.

Since one-element sets are closed, this restriction is in accordance with the above-mentioned need to consider one-element sets. (In a more general framework of a LCHS space, the requirement that all one-point sets be closed is one of the reasons why we impose the restriction that the topology must be Hausdorff: for Hausdorff topological spaces, this requirement is satisfied.)

With this restriction, the set $\mathcal{C}$ of possible outcomes is the class of all closed subsets of the space $\mathbb{R}^d$; this class is usually denoted by $\mathcal{F}(\mathbb{R}^d)$, or simply $\mathcal{F}$, for short.

It is worth mentioning that the decision to restrict ourselves to closed sets was made already in the pioneering book (Mathéron, 1975) on random sets.

We need a topology on $\mathcal{F}$. To finalize our definition of closed random sets, we must specify a σ-field on $\mathcal{F}$. To specify such a field, we will follow the same idea as with random vectors – namely, we will define a topology on $\mathcal{F}$ and then consider the σ-field of all the Borel sets in this topology (i.e., the smallest σ-filed that contains all sets which are open or closed in the sense of this topology.

In his monograph (Mathéron, 1975), Mathéron described a natural topology that he called a *hit-or-miss* topology. (It is worth mentioning that this topology was first introduced in a completely different problem: the construction of the regularized dual space of a C*-algebra (Fell, 1961), (Fell, 1962).)

Mathéron's motivations and definitions work well for random sets. However, they are difficult to directly generalize to random fuzzy sets. Since the main objective of this paper is to

describe and analyze random fuzzy sets, we will present a different derivation of this topology, a derivation that can be naturally generalized to random fuzzy sets.

This alternative definition is based on the fact that on the class of closed sets, there is a natural order $A \supseteq B$. The meaning of this order is straightforward. If we have a (closed) set A of possible vectors, this means that we only have partial information about the vector. When we gain additional information, this enables us to reduce the original set A to its subset B. Thus, $A \supseteq B$ means that the set B carries more information about the (unknown) vector than the set A.

It turns out that in many important situations, this order enables us to describe a natural topology on the corresponding ordered set. Specifically, this topology exists when the corresponding order forms a so-called continuous lattice. To describe this topology in in precise terms, we will thus need to recall the basic notions of lattice theory and the definition of a continuous lattice.

Basics of lattice theory. Lattices are a particular case of *partially ordered sets* (also knows as *posets*); so, to define the lattice, we must first recall the definition of a poset. A poset is defined as a set X together with a relation $\leq$ which satisfies the following properties:

(i) $\leq$ is *reflexive*, i.e., $x \leq x$ for all $x \in X$;

(ii) $\leq$ is *anti-symmetric*, i.e., for all $x, y \in X$, $x \leq y$ and $y \leq x$ imply that $x = y$;

(iii) $\leq$ is *transitive*, i.e., for all $x, y, z \in X$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

An element $u \in X$ is called an *upper bound* for a pair x, y is $x \leq u$ and $y \leq u$. Dually, an element $v \in X$ is called a *lower bound* for a pair x, y is $v \leq x$ and $v \leq y$. These concepts can be naturally extended to arbitrary collections of elements $A \subseteq X$: an element $u \in X$ is an *upper bound* for A if $x \leq u$ for all $x \in A$; an element $v \in X$ is a *lower bound* for A if $v \leq x$ for all $x \in A$.

The least upper bound is exactly what it sounds like: the least of all the upper bounds. Similarly, the greatest lower bound is the greatest of all the lower bounds. In precise terms, an element $u \in X$ is called the *least upper bound* of the set A if it satisfies the following two conditions:

a) u is an upper bound for the set A (i.e., $x \leq u$ for all $x \in A$); and

b) if w is an upper bound for the set A, then $u \leq w$.

The least upper bound of a set A is also called its *join* or *supremum* and is usually denoted by $\wedge A$ or $\sup A$.

Similarly, an element $v \in X$ is called the *greatest lower bound* of the set A if it satisfies the following two conditions:

a) v is a lower bound for the set A (i.e., $v \leq x$ for all $x \in A$); and

b) if w is an lower bound for the set A, then $w \le v$.

The greatest lower bound of a set A is also called its *meet* or *infimum* and is usually denoted by $\vee A$ or $\inf A$.

For two-element sets, $\wedge\{a, b\}$ is usually denoted by $a \wedge b$ and $\vee\{a, b\}$ by $a \vee b$.

A poset X is called a *lattice* if $a \wedge b$ and $a \vee b$ exist for all pairs $a, b \in X$. A poset X is called a *complete lattice* if both $\vee A$ and $\wedge A$ exist for any set $A \subseteq X$.

Examples of lattices. The set $\mathbb{R}$ of all real numbers with a standard order is a lattice, with $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. This set is not a complete lattice, since it does not have a join $\vee\mathbb{R}$: no real number is larger than all the others.

A slight modification of this example makes it a complete lattice: namely, the set of all real numbers from an interval $[\underline{x}, \overline{x}]$ is a complete lattice.

The set of all n-dimensional vectors $a = (a_1, \ldots, a_n)$ with a component-wise order

$$a \le b \leftrightarrow a_1 \le b_1 \,\&\, \ldots \,\&\, a_n \le b_n$$

is also a lattice, with $(a \vee b)_i = \max(a_i, b_i)$ and $(a \wedge b)_i = \min(a_i, b_i)$. Another example of a lattice is the set of all the functions, with $(f \vee g)(x) = \max(f(x), g(x))$ and $(f \wedge g)(x) = \min(f(x), g(x))$.

Continuous lattices and Lawson topology. In some posets, in addition to the original relation $\le$ ("smaller"), we can define a new relation $\ll$ whose meaning is "much smaller". This relation is called *way below*.

The formal definition of $\ll$ requires two auxiliary notions: of a directed set and of a dcpo. A set $A \subseteq X$ is called *directed* if every finite subset of A has an upper bound in A. Of course, in a complete lattice, every set is directed.

A poset X is called a *directed complete partial order (dcpo)*, if each of its directed subsets D has a supremum $\vee D$. In particular, every complete lattice is a dcpo.

We say that x is *way below* y ($x \ll y$) if for every directed subset D for which $y \le \vee D$, there exists an element $d \in D$ such that $x \le d$.

In particular, a 1-element set $D = \{y\}$ is always directed, and for this set, we conclude that $x \le y$ – i.e., that "way below" indeed implies $\le$. Another simple example: for a natural order on the set of real numbers, $x \ll y$ simply means $x < y$.

From the common sense viewpoint, we expect that if x is way below y and z is even smaller than x, then z should also be way below y. This is indeed true for the above definition: indeed, if $x \ll y$ and $z \le x$, then $z \ll y$, then for every D, we have $x \le d$ for some d and hence $z \le x \le d$, i.e., $z \le d$ for that same element $d \in D$. Similarly, we can proof all three statements from the following lemma.

Lemma 1. *Let $(L, \le)$ be a dcpo. Then, for any u, x, y, z in L, we have:*

(i) $x \ll y$ *implies* $x \le y$;

(ii) $u \le x \ll y \le z$ *implies* $u \ll z$;

(iii) if $x \ll z$ *and* $y \ll z$, *then* $x \vee y \ll z$.

For example, to prove (iii), for every directed set D, we must prove that $z \le \vee D$ implies that $x \vee y \le d$ for some $d \in D$. Indeed, from $x \ll z$ and $y \ll z$, we conclude that $x \le d_x$ and $y \le d_y$ for some $d_x, d_y \in D$. Since D is a directed set, there exists an element $d \in D$ which is an upper bound for both d_x and d_y, i.e., for which $d_x \le d$ and $d_y \le d$. From $x \le d_x$ and $d_x \le d$, we conclude that $x \le d$ and similarly, that $y \le d$, so d is an upper bound for x and y. By definition of the least upper bound $x \vee y$, it must be smaller than or equal than any other upper bound, hence $x \vee y \le d$. The statement is proven.

We can define a topology if we take, as a subbase, sets $\{y \in X : y \ll x\}$ and $\{y \in X : x \not\le y\}$ for all $x \in X$. In other words, as open sets for this topology, we take arbitrary unions of finite intersections of these sets. This topology is called a *Lawson topology*.

It is worth mentioning that for the standard order on real numbers, the sets $\{y \in X : y \ll x\} = \{y : y < x\}$ and $\{y \in X : x \not\le y\} = \{y : y > x\}$ are indeed open, and the corresponding topology coincides with the standard topology on the real line.

There is an important reasonably general case when the Lawson topology has useful properties: the case of a continuous lattice. A complete lattice X is called a *continuous lattice* if every element $x \in X$ is equal to the union of all the elements which are way below it, i.e., if $x = \vee\{y \in X : y \ll x\}$ for all $x \in X$. It is known that on every continuous lattice $(X, \le)$, the Lawson topology is *compact* and *Hausdorff*; see, e.g., (Gierz et al., 1980).

Final definition of a (closed) random set and the need for further analysis. In our analysis of random sets, we will use the Lawson topology to describe the σ-algebra of subsets of $\mathcal{F}$ – as the class of all Borel sets in the sense of this topology.

From the purely mathematical viewpoint, this is a (somewhat abstract but) perfect definition. However, since our objective is to make this definition applicable to practical problems, we need to first reformulate this general abstract definition in more understandable closer-to-practice terms.

THE LAWSON TOPOLOGY OF CLOSED SETS

First try. Let us describe what these notions lead to in the case of closed sets. For the class $\mathcal{F}$ of closed sets, there is a natural ordering relation $\le$: a set inclusion $F \subseteq F'$. The corresponding poset $(\mathcal{F}, \subseteq)$ is indeed a complete lattice: indeed, for every family F_i $(i \in I)$ of closed sets, there exist both the infimum and the supremum:

$$\wedge\{F_i \in \mathcal{F} : i \in I\} = \bigcap\{F_i : i \in I\};$$

$$\vee\{F_i : i \in I\} = \text{the closure of } \bigcup\{F_i : i \in I\}.$$

The resulting complete lattice is not continuous. Let us show that with $\subseteq$ as the ordering relation $\le$, $\mathcal{F}$ is *not* continuous. Specifically, as we will show, that, for example, in the 1-dimensional case $\mathbb{R}^d = \mathbb{R}$, the definition $F = \vee\{G \in \mathcal{F} : G \ll F\}$ of a continuous lattice is violated for $F = \mathbb{R}$.

For that, we will prove that for every two sets F and G, $G \ll F$ implies that $G = \emptyset$. We will prove this by reduction to a contradiction. Let us assume that $G \neq \emptyset$ and $G \ll F$. Since the set G is not empty, it contains an element $x \in G$. For this element x, the one-point set $S = \{x\}$ is a closed subset of G: $S \subseteq G$. By our first-try definition of $\le$, this means that $S \le G$. We have already mentioned that if $S \le G$ and $G \ll F$, then $S \ll F$. By definition of the "way below" relation $\ll$, this means that if $F \le \vee D$, then there exists an element $d \in D$ for which $S \le d$. In particular, we can take as D the family $\{d_n\}_n$, where for every positive integer n, $d_n \stackrel{\text{def}}{=} \left(-\infty, x - \frac{1}{n}\right] \cup \left[x + \frac{1}{n}, +\infty\right)$. It is easy to see that $d_{n_1} \vee \ldots \vee d_{n_k} = d_{\max(n_1,\ldots,n_k)}$, so the family D is indeed directed. The union $\bigcup\limits_n d_n$ of these sets is equal to $(-\infty, x) \cup (x, +\infty)$. Thus, the closure $\vee D$ of this union is the entire real line $\mathbb{R}$. Since F is a subset of the real line, we have $F \le \vee D$. However, for every $d_n \in D$, we have $x \notin d_n$ and thus, $S \not\le d_n$. The contradiction shows that a non-empty set G cannot be way below any other set F. Therefore, the only set G for which $G \ll \mathbb{R}$ is an empty set, so $\vee\{G \in \mathcal{F} : G \ll \mathbb{R}\} = \emptyset \neq \mathbb{R}$.

Correct definition. Fortunately, a small modification of the above definition makes $\mathcal{F}$ a continuous lattice. Namely, as the desired ordering relation $\le$ on the class $\mathcal{F}$ of all closed sets, we can consider the reverse inclusion relation $\supseteq$.

It is easy to show that $(\mathcal{F}, \supseteq)$ is a complete lattice: indeed,

$$\wedge\{F_i \in \mathcal{F} : i \in I\} = \text{the closure of } \bigcup\{F_i : i \in I\};$$

$$\vee\{F_i : i \in I\} = \bigcap\{F_i : i \in I\}.$$

Let us prove that $(\mathcal{F}, \supseteq)$ is a continuous lattice. By definition, a continuous lattice means that for every closed set $F \in \mathcal{F}$, we have $F = \vee\{G \in \mathcal{F} : G \ll F\}$ for every set $F \in \mathcal{F}$. Since $G \ll F$ implies $G \le F$, i.e., $G \supseteq F$, we thus have $\vee\{G \in \mathcal{F} : G \ll F\} \supseteq F$. So, to prove the desired equality, it is sufficient to prove that $F \supseteq \vee\{G \in \mathcal{F} : G \ll F\}$.

We have already mentioned that in the lattice $(\mathcal{F}, \supseteq)$, the union $\vee$ is simply the intersection of the corresponding sets, so the desired property can be rewritten as $F \supseteq \cap\{G \in \mathcal{F} : G \ll F\}$. It turns out that it is easier to prove the equivalent inclusion of complements, i.e., to prove that $F^c \subseteq \cup\{G^c : G \in \mathcal{F}, G \ll F\}$ (where F^c denotes the complement of the set F).

There is no easy and intuitive way to immediately prove this result, because the notion of "way below" is complex and is therefore not intuitively clear. So, to be able to prove results about this relation $\ll$, let us reformulate it in an equivalent more intuitive way.

Lemma 2. *For $X = \mathbb{R}^d$ (and, more generally, for an arbitrary locally compact Hausdorff second countable topological space X), for $F, G \in \mathcal{F}(X)$, $F \ll G$ if and only if there exists a compact set $K \subseteq X$ such that $F^c \subseteq K \subseteq G^c$.*

Proof.

(i) **Sufficiency.** Let $F, G \in \mathcal{F}(X)$ and let K be a compact set such that $F^c \subseteq K \subseteq G^c$. Let $G \le \vee D$ for some directed family D, i.e., let $G \supseteq \vee D$. We already know that $\vee$ is simply an intersection, i.e., $G \supseteq \cap D$. In terms of complements, we get an equivalent inclusion $G^c \subseteq \cup\{d^c : d \in D\}$. Since $K \subseteq G^c$, we conclude that $K \subseteq \cup\{d^c : d \in D\}$. Complements d^c to closed sets $d \in D$ are open sets. Thus, the compact K is covered by a family of open sets d^c, $d \in D$. By definition of a compact set, this means that we can select a finite subcover, i.e., conclude that $K \subseteq F_1^c \cup \ldots \cup F_n^c$ for some closed sets $F_i \in D$. Since $F^c \subseteq K$, we thus have $F^c \subseteq F_1^c \cup \ldots \cup F_n^c$, i.e., equivalently, $F \supseteq F_1 \cap \ldots \cap F_n$, i.e., $F \le F_1 \vee \ldots \vee F_n$.

Since sets $F_1, \ldots, F_n$ belong to the directed family D, this family must also contain an upper bound $d \in D$. By definition of the least upper bound, we have $F_1 \vee \ldots \vee F_n \le d$, hence $F \le d$. Thus, indeed, $G \ll F$.

(ii) **Necessity.** Let $F \ll G$. Since the underlying topological space X is locally compact, each point $x \in G^c$ has a compact neighborhood $Q_x \subseteq G^c$ such that its interior $\overset{\circ}{Q}_x$ contains x. We thus have $G^c = \cup\{\overset{\circ}{Q}_x : x \in G^c\}$ or, equivalently,

$$G = \cap\{(\overset{\circ}{Q}_x)^c : x \in G^c\} = \vee\{(\overset{\circ}{Q}_x)^c : x \in G^c\}.$$

Finite unions of closed sets $(\overset{\circ}{Q}_x)^c$ form a directed family D, for which the union is the same set G. Since $G \le \vee D$ and $F \ll G$, we conclude (by definition of the "way below" relation) that $F \le \vee\{(\overset{\circ}{Q}_{x_i})^c : i = 1, \ldots, n\}$, $x_i \in G^c$. Thus, $F \supseteq \bigcap\limits_{i=1}^{n} (\overset{\circ}{Q}_{x_i})^c$ or, equivalently, $F^c \subseteq \bigcup\limits_{i=1}^{n} (\overset{\circ}{Q}_{x_i})$. Therefore, $F^c \subseteq \bigcup\limits_{i=1}^{n} Q_{x_i} \subseteq G^c$. Since each set Q_{x_i} is compact, their union is also compact, so we have the desired inclusion $F^c \subseteq K \subseteq G^c$ with $K = \bigcup\limits_{i=1}^{n} Q_{x_i}$. The lemma is proven.

Proof that $(\mathcal{F}(X), \supseteq)$ is a continuous lattice. Let us now use this Lemma 2 to show that $(\mathcal{F}(X), \supseteq)$ is a continuous lattice. We have already mentioned that to prove this fact, we must prove that $F^c \subseteq \cup\{G^c : G \in \mathcal{F}, G \ll F\}$ for every closed set $F \in \mathcal{F}(X)$. Indeed, let F be a closed set. Since X is locally compact, for every point x from the open set F^c, there exists a compact neighborhood $K_x \subseteq F^c$ such that its interior $\overset{\circ}{K}_x$ contains x. The complement $A = (\overset{\circ}{K}_x)^c$ to this (open) interior is a closed set $A \in \mathcal{F}(X)$, for which $x \in A^c = \overset{\circ}{K}_x \subseteq K_x \subseteq F^c$. So, due to Lemma 2, $A \ll F$. In other words, if $x \in F^c$, then there is an $A \in \mathcal{F}$ such that $x \in A^c$ and $A \ll F$. So, we conclude that

$$F^c \subseteq \cup\{G^c : G \in \mathcal{F}, G \ll F\},$$

i.e., that $(\mathcal{F}(X), \supseteq)$ is indeed a continuous lattice.

Lawson topology on the class of all closed sets. Since $(\mathcal{F}(X), \supseteq)$ is a continuous lattice, we can define Lawson topology for this lattice. For this lattice, let us reformulate

the general abstract notion of the Lawson topology in more understandable terms. In the following text, we will denote the class of all compact subsets of the space X by $\mathcal{K}(X)$, and the class of all open subsets of X by $\mathcal{O}(X)$. When the space X is clear from the context, we will simply write $\mathcal{K}$ and $\mathcal{O}$.

Theorem 1. *For every LCHS X, the Lawson topology on the continuous lattice $(\mathcal{F}(X), \supseteq)$ is generated by the subbase consisting of subsets of $\mathcal{F}$ of the form $\mathcal{F}^K \stackrel{\text{def}}{=} \{F \in \mathcal{F} : F \cap K = \emptyset\}$ and $\mathcal{F}_G \stackrel{\text{def}}{=} \{F \in \mathcal{F} : F \cap G \neq \emptyset\}$, where $K \in \mathcal{K}$ and $G \in \mathcal{O}$.*

Comment. The class $\{F \in \mathcal{F} : F \cap K = \emptyset\}$ is the class of all random sets F which *miss* the set K, and the class $\{F \in \mathcal{F} : F \cap G \neq \emptyset\}$ is the class of all random sets F which *hit* the set G. Because of this interpretation, the above topology is called the hit-or-miss topology (Mathéron, 1975). So, the meaning of Theorem 1 is that the Lawson topology on $(\mathcal{F}(X), \supseteq)$ coincides with the hit-or-miss topology.

Proof. By definition, the Lawson topology has a subbase consisting of sets of the form $\{F' \in \mathcal{F} : F \ll F'\}$ and $\{F' \in \mathcal{F} : F \not\leq F'\}$ for all $F \in \mathcal{F}$, where $\leq$ is $\supseteq$. It turns out that to prove the theorem, it is sufficient to reformulate these sets in terms of $\supseteq$.

(i) Clearly,

$$\{F' \in \mathcal{F} : F \not\leq F'\} = \{F' \in \mathcal{F} : F' \cap F^c \neq \emptyset\}.$$

(ii) For $K \in \mathcal{K}$, we have

$$\{F \in \mathcal{F} : F \cap K = \emptyset\} = \bigcup_{F' \in \mathcal{F}, F' \subseteq K^c} \{F \in \mathcal{F} : F' \ll F\}.$$

Indeed, if $F \in \mathcal{F}$ and $F \cap K = \emptyset$, then K is a subset of the open set F^c. Since the space X is locally compact, there exists an open set B and a compact K' such that $K \subseteq B \subseteq K' \subseteq F^c$. The complement $G = B^c$ to an open set B is a closed set for which $K \subseteq G^c \subseteq K' \subseteq F^c$. By Lemma 2, this means that $G \ll F$ and, of course, $G \subseteq K^c$.

Conversely, let $f \in \mathcal{F}$ and $F' \ll F$ with $F' \subseteq K^c$. Then, by the same Lemma 2, there is a compact set K' with $(F')^c \subseteq K' \subseteq F^c$. On the other hand, $F' \subseteq K^c$ implies that $K \subseteq (F')^c$, so we have $K \subseteq F^c$ and $K \cap F = \emptyset$. The theorem is proven.

From topology to metric: need and possibility. We have defined topology on the class of all closed sets. Topology describes the intuitive notion of closeness in *qualitative* terms. From the viewpoint of applications, it is convenient to use *quantitative* measures of closeness such as a *metric*.

Before we start describing a corresponding metric, let us first prove that this metric is indeed possible, i.e., that the corresponding topological space is metrizable. It is known that for every continuous lattice, the corresponding Lawson topology is compact and Hausdorff. Let us prove that for LCHS spaces X, the Lawson topology on $(\mathcal{F}(X), \supseteq)$ is not only compact and Hausdorff, it is also second countable – and therefore metrizable.

Theorem 2. *For every LCHS X, the Lawson topology on $(\mathcal{F}(X), \supseteq)$ is second countable.*

Proof. This proof is given in (Mathéron, 1975) for the hit-or-miss topology. We reproduce it here for completeness.

Since X is locally compact Hausdorff second countable space, there exists a countable basis $\mathcal{B}$ for the topology $\mathcal{O}$ consisting of *relatively compact* sets $B \in \mathcal{B}$ (i.e., sets whose closure $\overline{B}$ is compact). The fact that $\mathcal{B}$ is a basis means that every open set $G \in \mathcal{O}$ can be represented as a union of open sets from this basis, i.e., that it can be represented in the form $G = \bigcup\limits_{i \in I} B_i$, where $B_i \in \mathcal{B}$, and $\overline{B}_i \subseteq G$.

By definition of the hit-or-miss topology, its subbase consists of the sets

$$\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \emptyset\}$$

for compact sets A and sets $\mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \emptyset\}$ for open sets A. Thus, the base of this topology consists of the finite intersections of such sets. The intersection of two sets $\mathcal{F}^A \stackrel{\text{def}}{=} \{F \in \mathcal{F} : F \cap A = \emptyset\}$ and $\mathcal{F}^{A'} \stackrel{\text{def}}{=} \{F \in \mathcal{F} : F \cap A' = \emptyset\}$ consists of all the sets F which do not have common points neither with A nor with A'; this is equivalent to saying that F has no common points with the union $A \cup A'$, i.e., that $\mathcal{F}^A \cap \mathcal{F}^{A'} = \mathcal{F}^{A \cup A'}$. Thus, finite intersections which form the basis of the hit-or-miss topology (or, equivalently, Lawson topology) have the form

$$\mathcal{F}^K_{G_1,\ldots,G_n} \stackrel{\text{def}}{=} \{F \in \mathcal{F} : F \cap K = \emptyset, F \cap G_i \neq \emptyset, i = 1, 2, \ldots, n\}$$

for all possible $K \in \mathcal{K}$ and $G_i \in \mathcal{O}$.

Let us consider the following subset of this basis:

$$\mathcal{T} = \left\{ \mathcal{F}^{\overline{D}_1 \cup \ldots \cup \overline{D}_k}_{B_1,\ldots,B_m} ; m, k \geq 0, B_i, D_j \in \mathcal{B} \right\}.$$

Clearly $\mathcal{T}$ is countable. We need to verify that $\mathcal{T}$ is a basis for the Lawson topology. For this, we need to prove that for every element $F \in \mathcal{F}$, in every open neighborhood of F, there is an open set from $\mathcal{T}$ that contains F. It is sufficient to prove this property only for the neighborhood from the known basis.

Indeed, let $F \in \mathcal{F}$ and let the open set $\mathcal{F}^K_{G_1,\ldots,G_n}$ be a neighborhood of F. We need to find a $U \in \mathcal{T}$ such that $F \in U \subseteq \mathcal{F}^K_{G_1,\ldots,G_n}$. We will prove this by considering two possible cases: $F = \emptyset$ and $F \neq \emptyset$.

If $F = \emptyset$, then we cannot have $F \cap G_i \neq \emptyset$, so n must be 0, and $\mathcal{F}^K_{G_1,\ldots,G_n} = \mathcal{F}^K$. Since $\mathcal{B}$ is a basis, the compact set K can be covered by open sets from this basis. Since K is compact, we can extract a finite cover from this cover, i.e., conclude that $K \subseteq D_1 \cup \ldots \cup D_k$ for some elements $D_j \in \mathcal{B}$. Thus, $K \subseteq \overline{D}_1 \cup \ldots \cup \overline{D}_k$. Clearly, the empty set has no common point with anyone, so $F \in \mathcal{F}^{\overline{D}_1 \cup \ldots \cup \overline{D}_k}$.

If a set has no common points with a larger set, then it will no common points with a subset either; so we conclude that $F \in \mathcal{F}^{\overline{D}_1 \cup \ldots \cup \overline{D}_k} \subseteq \mathcal{F}^k$. So, we can take $U = \mathcal{F}^{\overline{D}_1 \cup \ldots \cup \overline{D}_k}$ as the desired neighborhood.

If $F \neq \emptyset$, then the fact that F belongs to the neighborhood $\mathcal{F}^K_{G_1,\ldots,G_n}$ means that $F \cap G \neq \emptyset$ for every i. This means that for every $i = 1, 2, \ldots, n$, we can pick a point $x_i \in F \cap G_i$. Since $\mathcal{B}$ is a basis, for every i, there exists an open set $B_i \in \mathcal{B}$ such that $x_i \in B_i \subseteq \overline{B}_i \subseteq G_i \cap K^c$.

This means that our set F not only has no common points with K, it also has no common points with the closed sets $\overline{B}_i$. Since the closed sets K and $F \cup \left(\bigcup\limits_{i=1}^{n} \overline{B}_i\right)$ are disjoint and the space X is Hausdorff, we can find two disjoint open sets A_1 and A_2 which contain these sets: $K \subseteq A_1$ and $F \cup \left(\bigcup\limits_{i=1}^{n} \overline{B}_i\right) \subseteq A_2$. Since $\mathcal{B}$ is a basis, we can represent the open set A_1 as a union of open sets from $\mathcal{B}$: $K \subseteq A_1 = \bigcup\limits_{j} D_j$ for some $D_j \in \mathcal{B}$ for which $\overline{D}_j \subseteq A_1$. The set K is compact and thus, from this open cover of K, we can extract a finite sub-cover, i.e., we have

$$K \subseteq \bigcup_{j=1}^{k} D_j \subseteq \bigcup_{j=1}^{k} \overline{D}_j \subseteq A_1 \subseteq A_2^c,$$

and $\left(\bigcup\limits_{j=1}^{k} \overline{D}_j\right) \cap A_2 = \emptyset$. Thus,

$$F \in U \stackrel{\text{def}}{=} \mathcal{F}^{\overline{D}_1 \cup \ldots \cup \overline{D}_k}_{B_1,\ldots,B_n} \subseteq \mathcal{F}^K_{G_1,\ldots,G_n}.$$

The theorem is proven.

METRICS ON CLOSED SETS

From theoretical possibility of a metric to a need for an explicit metric. In the previous section, we have shown that for every LCHS X (in particular, for $X = \mathbb{R}^d$), when we define Lawson topology on the set $\mathcal{F}(X)$ of all closed subsets of X, then this set $\mathcal{F}(X)$ becomes metrizable. This means that *in principle*, this topology can be generated by a metric.

However, from the *application* viewpoint, it is not enough to consider a theoretical possibility of a metric, it is desirable to describe a specific explicit example of a metric.

Known metric on the class of all closed sets: Hausdorff metric. Intuitively, the metric describes the notion of a distance between two sets. Before we start describing an explicit metric which is compatible with the Lawson topology, let us first describe natural ways to describe such a distance.

For points $x, y \in \mathbb{R}^d$, the standard distance $\delta(x, y)$ can be described in terms of neighborhoods: the distance is the smallest value $\varepsilon > 0$ such that x belongs to the ε-neighborhood of y and y is in the ε-neighborhood of x.

It is natural to extend the notion of ε-neighborhood from points to sets. Namely, a set A is collection of all its points; thus, an ε-neighborhood $N_\varepsilon(A)$ of a set A is a collection of ε-neighborhoods of all its points. In other words, we can define

$$N_\varepsilon(A) \stackrel{\text{def}}{=} \{x \in X : \delta(x, a) \leq \varepsilon \text{ for some } a \in A\}.$$

Now, we can define the distance $d_H(A, B)$ between the two sets as the smallest smallest value $\varepsilon > 0$ for which A is contained in the ε-neighborhood of B and B is contained in the ε-neighborhood of A:

$$H_\delta(A, B) \stackrel{\text{def}}{=} \inf\{\varepsilon > 0 : A \subseteq N_\varepsilon(B) \,\&\, B \subseteq N_\varepsilon(A)\}.$$

This metric is called the *Hausdroff metric* (after the mathematician who proposed this definition).

Examples. Hausdorff distance between a one-point set $\{0.3\}$ and an interval $[0, 1]$ is 0.7: indeed, all the elements from the set $\{0.3\}$ are contained in the interval $[0, 1]$, and every element of the interval $[0, 1]$ is contained in the 0.7-vicinity of the point 0.3.

This example can be viewed as a degenerate case of computing the Hausdorff distance between two intervals $[\underline{a}, \overline{a}]$ and $[\underline{b}, \overline{b}]$ – namely, the case when $\underline{a} = \overline{a}$. In general, the Hausdorff distance between the two intervals is equal to $\max(|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|)$.

Limitations of the known metric. Hausdorff metric works well for bounded sets in $\mathbb{R}^d$ or, more generally, for subsets of a compact set. However, if we use the Haudorff metric to described distances between arbitrary closed sets $A, B \subseteq \mathbb{R}^d$, we often get meaningless infinities: for example, in the plane, the Hausdorff distance between any two non-parallel lines is infinity.

How to overcome these limitations: the notion of compactification. To overcome the above limitation, it is desirable to modify the definition of a Hausdorff metric.

Since Hausdorff metric works well for compact spaces, a natural idea is to somehow embed the original topological spaces into a compact set. Such a procedure is known as *compactification*.

Simplest compactification. In the simplest compactification, known as *Alexandroff* (or *one-point*) *compactification*, we simply add a single point ∞ to the original space X.

The corresponding topology on $X \cup \{\infty\}$ is defined as follows:

- if a set $S \subseteq G$ does not contain the new point ∞, then it is open if and only if it was open in the original topology;
- if a set S contains ∞, then it is called open if and only if its complement S^c is a compact subset of the original space X.

From points to closed sets. This compactification can be extended to an abritrary closed set F: namely, a set F which was closed in X is not necessarily closed in the new space $X \cup \{\infty\}$, so we have to take the closure $\overline{F}$ of this set.

If we have a matric δ' on the compactification, then we can define a distance between closed sets $F, F' \in \mathcal{F}$ as $H_{\delta'}(\overline{F}, \overline{F'})$.

Compactification of $\mathbb{R}^d$. For $\mathbb{R}^d$, the one-point compactification can be implemented in a very natural way, via a so-called *stereographic projection*. Specifically, we can interpret

a one-point compactification of the space $\mathbb{R}^d$ as a sphere $\mathbb{S}^d \subseteq \mathbb{R}^{d+1}$. The correspondence between the original points of $\mathbb{R}^d$ and $\mathbb{S}^d$ is arranged as follows. We place the space $\mathbb{R}^d$ horizontally, and we place the sphere $\mathbb{S}^d$ on top of this plane, so that its "South pole" is on that plane. Then, to find the image $\pi(x)$ of a point $x \in \mathbb{R}^d$ on the sphere, we connect this point x with the "North pole" of the sphere by a straight line, and take, as $\pi(x)$, the intersection between this line and the sphere. In this manner, we cover all the points on the sphere except for the North pole itself. The North pole corresponds to infinity – in the sense that if $x \to \infty$, then $\pi(x)$ tends to the North pole.

In this sense, $\mathbb{S}^d$ is a one-point compactification of the original space $\mathbb{R}^d$: it is a compact space, and it is obtained by adding a point (North pole) to (the image of) $\mathbb{R}^d$.

From points to closed sets. By using this construction, we can naturally find the projection of an arbitrary closed set $F \in \mathcal{F}(\mathbb{R}^d)$ as the collection of all the projections of all its points, i.e., as $\pi(F) = \{\pi(x) : x \in F\}$. The only minor problem is even while we started with a closed set F, this collection $\pi(F)$, by itself, may be not closed. For example, a straight line on a plane is closed, but its projection is closed – since it has point arbitrary close to the North pole but not the North pole itself.

To resolve this problem, we can then take a closure of this image. Thus, we arrive at the following *stereographic Hausdorff metric* $H_{\delta'}(\overline{\pi(F)}, \overline{\pi(F')})$, where δ' is the standard distance on the sphere $\mathbb{S}^d$.

Relation between this metric and the Lawson topology. It turns out that the Lawson topology on $\mathcal{F}(\mathbb{R}^d)$ is compatible with the Hausdorff metric on a one-point compactification.

For the stereographic Hausdorff metric, this equivalence is proven, e.g., in (Rockafeller and Wets, 1984); for the general LCHS space, this is proven in (Wei and Wang, in press).

RANDOM CLOSED SETS: FINAL DEFINITION AND CHOQUET THEOREM

Final definition of a random (closed) set. With the material developed in previous sections, we are now ready to formulate rigorously the concept of random closed sets as *bona fide* random elements. These are generalizations of random vectors and serve as mathematical models for observation processes in which observables are sets rather than points.

Again, let $\mathcal{F}(\mathbb{R}^d)$ ($\mathcal{F}(X)$) be the space of closed sets of $\mathbb{R}^d$, or more generally, of a LCHS space X. Let $\sigma(\mathcal{F})$ denote the Borel σ-field of subsets of $\mathcal{F}$ where $\mathcal{F}$ is equipped with the Lawson topology. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A map $V : \Omega \to \mathcal{F}$, which is $\mathcal{A}$-$\sigma(\mathcal{F})$-measurable, is called a *random closed set* (RCS) on $\mathbb{R}^d$. As usual, the probability law governing V is the probability measure $P_V = PV^{-1}$ on $\sigma(\mathcal{F})$.

Application-motivated need to describe a random set. A random set is, in effect, a probability measure on the class of all closed subsets $\mathcal{F}(X)$ of the LCHS X. In principle, we can thus describe the random set by listing, for different subsets $S \subseteq \mathcal{F}$, the corresponding probability $P(S)$.

From the application viewpoint, however, this description is duplicating – e.g., since due

to additivity, the probability $P_V(S \cup S')$ of a union of two disjoint sets is equal to the sum $P_V(S)+P_V(S')$ of the probabilities $P_V(S)$ and $P_V(S')$ and thus, does not have to be described independently.

For random numbers and for random vectors, the solution was to only give probability of sets from a given subbase. For real numbers, this subbase consisted of sets $(-\infty, x]$ – which let to the notion of a probability distribution. For elements of $\mathbb{R}^d$, we considered sets of the type $(-\infty, x_1] \times \ldots \times (-\infty, x_d]$. For the hit-or-miss topology on the space $\mathcal{F}(X)$, the subbase consists of the sets $\{F : F \cap A = \emptyset\}$ for compact A and $\{F : F \cap A \neq \emptyset\}$ for open sets A. It is therefore reasonable to define the probability only on the sets of these type.

The set $\{F : F \cap A = \emptyset\}$ is a complement to the set $\{F : F \cap A \neq \emptyset\}$, so we do not have to describe its probability separately: it is sufficient to describe the probability of the sets $\{F : F \cap A \neq \emptyset\}$.

Resulting description: capacity functionals. As a result, we arrive at the following definition. Once a random closed set X is defined, we can define the mapping $T : \mathcal{K} \to [0, 1]$ as $T(K) = P\{\omega : X(\omega) \cap K \neq \emptyset\} = P_X(\mathcal{F}_K)$. This mapping is called a *capacity functional* of a random closed set.

Is this description sufficient? A natural question is whether from this functional we can uniquely determined the corresponding probability measure. A related question is what are the conditions under which such a measure is possible.

It can be checked that T satisfies the following properties:

(i) $0 \le T(\cdot) \le 1$, $T(\emptyset) = 0$;

(ii) If $K_n \searrow K$ then $T(K_n) \searrow T(K)$;

(iii) T is *alternating of infinite order*, i.e., T is monotone (with respect to $\subseteq$) and for $K_1, K_2, \ldots, K_n$, $n \ge 2$,

$$T\left(\bigcap_{i=1}^{n} K_i\right) \le \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} T\left(\bigcup_{i \in I} K_i\right),$$

where $|I|$ denotes the cardinality of the set I.

It turns out that under these conditions, there is indeed such a measure.

Theorem (Choquet) *Let $T : \mathcal{K} \to \mathbb{R}$. Then the following two statements are equivalent to each other:*

- *there exists a probability measure Q on $\sigma(\mathcal{F})$ for which $Q(\mathcal{F}_K) = T(K)$ for all $K \in \mathcal{K}$;*
- *T satisfies the conditions (i), (ii) and (iii).*

If one of these statements is satisfied, then the corresponding probability measure Q is uniquely determined by T.

For a proof, see (Mathéron, 1975).

This theorem is the counter-part of the Lebesgue-Stieltjes theorem characterizing probability measures on the Borel σ-field of $\mathbb{R}^d$ in terms of multivariate distribution function S of random vectors. For subsequent developments of RCS, see e.g. (Nguyen, 2006).

FROM RANDOM SETS TO RANDOM FUZZY SETS

Need for fuzziness. As stated in the Introduction, random set observations are coarse data containing the true, unobservable outcomes of random experiments on phenomena.

A more general type of random and imprecise observations occurs when we have to use *natural language* to describe our perception. For example, the risk is "high" is an "observation" containing also imprecision at at a higher level due to the *fuzziness* in our natural language. Modeling fuzziness in natural language, i.e., modeling the meaning of terms is crucial if we wish to process information of this type.

How we can describe fuzziness. Following Zadeh, we use the theory of *fuzzy sets* to model the meaning of a nature language; see, e.g., (Nguyen and Walker, 2006) for fuzzy sets and fuzzy logics.

The theory is in fact valid for any LCHS space X. For concreteness, one can keep in mind an example $X = \mathbb{R}^d$. By a *fuzzy subset* of X we mean a function $f : X \to [0, 1]$ where $f(x)$ is the degree to which the element x is compatible with the fuzzy concept represented by f. This function f is also called a *membership function*.

For example, the membership function f of the fuzzy concept $A =$ "small non-negative numbers" of $\mathbb{R}$ could be $f(x) = 0$, if $x < 0, e^{-x}$ for $x \geq 0$, where $f(x) = e^{-x}$ is the degree to which x is considered as a "small non-negative number".

Ordinary subsets of X are special cases of fuzzy subsets of X, where for $A \subseteq X$, its indicator function $I_A : \mathbb{R}^d \to \{0, 1\} \subseteq [0, 1]$, $I_A(x) = 1$ if $x \in A$, and $I_A(x) = 0$ if $x \notin A$, is the membership function of A.

An alternative way to describe fuzzy sets: α-cuts. Informally, the fact that the actual (unknown) value $_{act} \in X$ satisfies the property described by a fully set f means the following:

- with confidence 1, the actual value x_{act} satisfies the condition $f(x_{act}) > 0$;
- with confidence 0.9, the actual value x_{act} satisfies the condition $f(x_{act}) > 1 - 0.9 = 0.1$, i.e., belongs to the set $\{x : f(x) \geq 0.1\}$;
- etc.

In view of this interpretation, instead of describing the fuzzy set by its membership function $f(x)$, we can alternatively describe it by the corresponding sets $A_\alpha \stackrel{\text{def}}{=} \{x \in X : f(x) \geq \alpha\}$ for different $\alpha \in [0, 1]$. These sets are called *alpha-cuts*.

Alpha-cuts are *nested* in the sense that if $\alpha < \alpha'$, then $A_\alpha \supseteq A_{\alpha'}$. So, a fuzzy set can be viewed as a nested family of sets (corresponding to different degrees of confidence); see, e.g., (Klir and Yuan, 1995), (Nguyen and Kreinovich, 1996), (Nguyen and Walker, 2006).

Fuzzy analog of closed sets. In our description of random sets, we limited ourselves to *closed* sets. Since a fuzzy set can be viewed as a nested family of sets (its alpha-cuts), it is reasonable to consider fuzzy sets in which all alpha-cuts are closed sets.

In other words, it is reasonable to restrict ourselves to fuzzy sets $f : X \to [0, 1]$ which have the following property: for every $\alpha \in \mathbb{R}$, the set $A_\alpha = \{x \in X : f(x) \geq \alpha\}$ is a closed subset of X. In mathematics, functions f with this property are called *upper semicontinuous* (*usc*, for short). We will thus call a fuzzy set a *fuzzy closed set* if its membership function is usc.

Closed sets are a particular case of fuzzy closed sets. We have already mentioned that traditional (crisp) sets are a particular case of fuzzy sets. It is easy to check that closed crisp sets are also closed as fuzzy sets.

Moreover, a subset of X is closed if and only if its indicator function is upper semicontinuous.

Towards a definition of a random fuzzy closed set. To formalize the concept of random fuzzy (closed) sets, we will therefore consider the space $USC(X)$ (where X is a LCHS space), of all usc functions on X with values in $[0, 1]$.

As we mentioned earlier, a natural way to define a notion of the random fuzzy closed set is to describe a topology on the set $USC(X)$. Similar to the case of closed sets, we will use the Lawson topology generated by a natural order on $USC(X)$.

The space $USC(X)$ has a natural pointwise order: $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \in X$. We can also consider the dual order $f \geq g$ (which is equivalent to $g \leq f$). We will now look for the corresponding Lawson topology!

THE CONTINUOUS LATTICE OF UPPER SEMICONTINUOUS FUNCTIONS

First try: $(USC(X), \leq)$. Let X as any LCHS space. By $USC(X)$ we mean the set of all usc functions $f : X \to [0, 1]$.

With the order relation $\leq$, $USC(X)$ is complete lattice, where $\left(\bigwedge_{j\in J} f_j\right)(x) = \inf\{f_j(x), j \in J\}$ and $\left(\bigvee_{j\in J} f_j\right)(x) = \sup\{\alpha \in [0, 1] : x \in A_\alpha\}$, where

$$A_\alpha = \text{ closure of } \bigcup_{j\in J} \{y \in X : f_j(y) \geq \alpha\}.$$

Remarks.

(i) Clearly $f = \inf\{f_j, j \in J, f_j \in USC(X)\} \in USC(X)$.

Indeed, let $\alpha \in [0,1]$, then $\{x \in X : f(x) < \alpha\} = \bigcup_{j \in J} \{x : f_j(x) < \alpha\}$ is an open set.

(ii) Let us explain why we cannot simply use pointwise supremum to define $\vee$.

Indeed, for $f_n(x) = I_{\left[\frac{1}{n}, +\infty\right)}(x), n \geq 1$, we have $f_n \in USC(X)$, but the pointwise supremum $\sup_n \left\{ I_{\left[\frac{1}{n}, +\infty\right)}(x), n \geq 1 \right\} = I_{(0,+\infty)}(x) \notin USC(X)$.

To define $\vee\{f_j : j \in J\}$, we proceed as follows.

For $\alpha \in [0,1]$, let

$$A_\alpha = \text{ closure of } \bigcup_{j \in J} \{y \in X : f_j(y) \geq \alpha\},$$

and define $f(x) = \sup\{\alpha \in [0,1] : x \in A_\alpha\}$. We claim that f is the least upper bound of $\{f_j : j \in J, f_j \in USC(X)\}$ in $USC(X)$ with respect to $\leq$.

Clearly each A_α is closed and $\alpha < \beta$ implies $A_\beta \subseteq A_\alpha$. As such $f \in USC(X)$.

Next, for $x \in X$, we have

$$x \in \text{ closure of } \bigcup_{j \in J} \{y : f_j(y) \geq f_i(x)\} = A_{f_i(x)}$$

for any $i \in J$. Thus, for every $i \in J$, we have $f_i(x) \leq f(x)$. Since this is true for every x, we thus conclude that $f \geq f_i$ for all $i \in J$.

Now, for $g \in USC(X)$, we can write $g(x) = \sup\{\alpha \in [0,1] : x \in A_\alpha(g)\}$ where $A_\alpha(g) = \{y \in X : g(y) \geq \alpha\}$.

For any $y \in \bigcup_{j \in J} \{x : f_j(x) \geq \alpha\}$, i.e., for any y for which $f_j(y) \geq \alpha$ for some j, we have $g(y) \geq \alpha$ if $g \geq f_i$ for all $i \in I$. Thus, $y \in A_\alpha(g)$, implying that $A_\alpha \subseteq A_\alpha(g)$, since $A_\alpha(g)$ is closed. But then $\sup\{\alpha \in [0,1] : x \in A_\alpha\} \leq \sup\{\alpha \in [0,1] : x \in A_\alpha(g)\}$ and hence $f \leq g$.

(iii) However, the complete lattice $(USC(X), \leq)$ is not continuous.

The proof of this statement is similar to the proof that the lattice $(\mathcal{F}(X))$ is not continuous. Indeed, let $f : X \to [0,1]$ be a function for which $f(x) = 1$ for all $x \in X$. Let us then show that the zero function, i.e. a function g for which $g(x) = 0$ for all $x \in X$, is the only function in $USC(X)$ which is way below f.

If suffice to show that for any real number $r > 0$, the function $r \cdot I_{\{y\}}(\cdot)$ (e.g., $\frac{1}{2} \cdot I_{\{0\}}(\cdot)$), is not way below f.

Let $f_n(x) = I_{(-\infty, y - \frac{1}{n}) \cup (y + \frac{1}{n}, +\infty)}$, then $\bigvee_{n \geq 1} f_n = 1$, but for any K, we have $\bigvee_{n=1}^{K} f_n(y) = 0$, thus $r \cdot I_{\{y\}} \not\leq \bigvee_{n=1}^{K} f_n$, implying that $r \cdot I_{\{y\}}$ is not way below f.

Correct description: $(USC, \geq)$**.** Thus, as in the case of $\mathcal{F}(X)$, we should consider the reverse order $\geq$.

Theorem 3. *The complete lattice* $(USC(X), \geq)$*, where*

$$\vee\{f_j : j \in J\} = \inf\{f_j : j \in J\}$$

and $\wedge\{f_j : j \in J\} = h$ *with* $h(x) = \sup\{\alpha \in [0,1] : x \in A_\alpha\}$ *and*

$$A_\alpha = \text{ closure of } \bigcup_{j \in J}\{y \in X : f_j(y) \geq \alpha\},$$

is continuous.

Before proving this theorem, we need a representation for elements of $USC(X)$. For every real number r and for every compact set $K \in \mathcal{K}(X)$, let us define an auxiliary function $g_{r,K}$ as follows: $g_{r,K}(x) = r$ if $x \in \overset{\circ}{K}$ and $g_{r,K}(x) = 1$ otherwise.

Lemma 3. *Any* $f \in USC(X)$ *can be written as*

$$f(\cdot) = \inf\{g_{r,K}(\cdot) : f(y) < r \text{ for all } y \in K\},$$

where infimum is taken over all pairs $r \in [0,1]$ *and* $K \in \mathcal{K}(X)$ *for which* $f(y) < r$ *for all* $y \in K$*.*

Comment. In this definition, the infimum of an empty set is assumed to be equal to 1.

Proof. Let $x \in X$. Let us consider two possible cases: $f(x) = 1$ and $f(x) < 1$.

(i) Let us first consider the case when $f(x) = 1$.

To prove the formula for this case, we will consider two cases: when $x \notin \overset{\circ}{K}$ and when $x \in \overset{\circ}{K}$.

In the first subcase, when $x \notin \overset{\circ}{K}$, we have $g_{r,K}(x) = 1$ by definition of the function $g_{r,K}$. Thus, the infimum is equal to 1, i.e., to $f(x)$.

In the second subcase, when $x \in \overset{\circ}{K}$, then there is no r such that $f(y) < r$ for all $y \in K$. Thus, the infimum is also equal to $1 = f(x)$.

(ii) Let us now consider the case when $f(x) < 1$.

We need to prove that $f(x)$ is the greatest lower bound of the values $g_{r,K}(x)$. Let us first prove that $f(x)$ is a lower bound. For that, we will again consider two subcases: $x \notin \overset{\circ}{K}$ and when $x \in \overset{\circ}{K}$.

When $x \in \overset{\circ}{K}$, we have $f(x) < r = g_{r,K}(x)$. When $x \notin \overset{\circ}{K}$, we have $f(x) < 1 = g_{r,K}(x)$. Thus, in both subcases, $f(x)$ is indeed a lower bound of $\{g_{r,K}(x)\}$.

Let us now prove that $f(x)$ is the greatest lower bound. In other words, let us prove that for any $\varepsilon > 0$, the value $f(x) + \varepsilon$ is not a lower bound of $\{g_{r,K}(x)\}$.

Indeed, let $r_0 = f(x) + \dfrac{\varepsilon}{2}$, then x belongs to the open set $\left\{y \in X : f(y) < r_0 = f(x) + \dfrac{\varepsilon}{2}\right\}$. By local compactness of X, we conclude that there is a compact set

$$K_0 \subseteq \{y : f(y) < r_0\}$$

such that $x \in \overset{\circ}{K}_0$, implying that $g_{r_0,K_0}(x) = r_0 < f(x) + \varepsilon$, i.e., that for every $y \in K_0$, we have $f(y) < r_0$ and $g_{r_0,K_0}(x) < f(x) + \varepsilon$. The Lemma is proven.

Now, we are ready to prove the theorem.

Proof of Theorem 3. Consider $(USC(X), \geq)$. For $f \in USC(X)$, we always have $f \leq \inf\{g \in USC(X) : g \ll f\}$, thus it is sufficient to show that

$$\inf\{g \in USC(X) : g \ll f\} \leq f = \inf\{g_{r,K} : f(y) < r \text{ for all } y \in K\}.$$

To prove this relation, it is sufficient to show that $g_{r,K} \ll f$ for any (r, K) such that $f(y) < r$ for all $y \in K$.

Indeed, let $F \subseteq USC(X)$ be a directed set such that $f \geq \vee F$, i.e., $\inf F \leq f$ (pointwise). To prove the desired "way below" relation, we need to find $h \in F$ such that $h \leq g_{r,K}$.

For (r, K) with $r > f \geq \inf F$ (pointwise on K), we will show that there exists an $h \in F$ such that $h < r$ on K. For any $h \in F$, denote

$$K_h \stackrel{\text{def}}{=} \{x \in K : r \leq h(x)\}.$$

Since h is usc, the set K_h is a closed set.

Let us prove that the intersection $\bigcap\limits_{h \in F} K_h$ of all these sets K_h is the empty set. Indeed, if $x \in \bigcap\limits_{h \in F} K_h$, then by definition of K_h, it means that $r \leq h(x)$ for all $h \in F$. This will imply that $r \leq \inf F$ on K, contradicting the fact that $r > \inf F$ on K.

Since every set K_h is closed, its complement K_h^c is open. From $\bigcap\limits_{h \in F} K_h = \emptyset$, we conclude that $\bigcup\limits_{h \in F} K_h^c = X$ and thus, $K \subseteq \bigcup\limits_{h \in F} K_h^c$. Since K is a compact set, from this open cover, we can extract a finite subcover, hence $K \subseteq \bigcup\limits_{i=1}^{n} K_{h_i}^c$ for some functions h_i. Thus, we have $\bigcap\limits_{i=1}^{n} K_{h_i} \subseteq K^c$. Since $K_h \subseteq K^c$ for all h, we thus conclude that

$$\bigcap_{i=1}^{n} K_{h_i} = \emptyset.$$

Since F is directed, we have

$$h \stackrel{\text{def}}{=} \vee(h_i : i = 1, \ldots, n) = \inf\{h_i : i = 1, \ldots, n\} \in F.$$

For this h, we have $K_h = \bigcap\limits_{i=1}^{n} K_{h_i} = \emptyset$. This means that for every $x \in K$, we have $h(x) < r$.

Now, for an arbitrary point $x \in X$, we have two possibilities: $x \in K$ and $x \notin K$. If $x \in K$, then $h(x) < r < g_{r,K}(x)$. On the other hand, if $x \notin K$, then $h(x) \le 1$ and $g_{r,K}(x) = 1$, hence also $h(x) \le g_{r,K}(x)$. So, for all $x \in X$, we have $h(x) \le g_{r,K}(x)$. The statement is proven.

Towards an application-oriented description of the corresponding Lawson topology. Since $(USC, \ge)$ is a continuous lattice, we can define the corresponding Lawson topology.

The Lawson topology is defined in very general, very abstract terms. To be able to efficiently apply this abstractly defined topology to random fuzzy closed sets, it is desirable to describe the Lawson topology for such sets in easier-to-understand terms. Such a description is given by the following result.

Theorem 4. *The following sets form a subbase for the Lawson topology on $(USC(X), \ge)$: the sets of the form*

$$\{f : f(y) < r \text{ for all } y \in K\}$$

for all $r \in (0, 1]$ and $K \in \mathcal{K}(X)$, and the sets

$$\{f : g(x) < f(x) \text{ for some } x \in X\}$$

for all $g \in USC(X)$.

Proof. By definition, the Lawson topology on $(USC(X), \ge)$ is generated by the subbase consisting of the sets $\{f : g \ll f\}$ and $\{f : g \not\ge f\}$. For the ordered set $(USC(X), \ge)$, clearly,

$$\{f : g \not\ge f\} = \{f : g(x) < f(x) \text{ for some } x \in X\}.$$

Let us now considers sets $\{f : g \ll f\}$. It is known that these sets form a base of a topology which is called *Scott topology*; see, e.g., (Gierz et al., 1980). A Scott open set is thus an arbitrary union of sets of the type $\{f : g \ll f\}$ with different g. So, to prove our theorem, it is sufficient to prove that the sets $\{f : f(y) < r \text{ for all } y \in K\}$ form a subbase of the Scott topology.

To prove this, we first prove that each such set is indeed a Scott open set; this is proved in the Lemma below. It then remains to verify that the above sets indeed form a subbase for the Scott topology.

Indeed, let A be a Scott open set which contains an element h. Since $A = \bigcup\limits_{g} \{f : g \ll f\}$, we have $h \in \{f : g \ll f\}$ for some g, i.e., $g \ll h$. It is easy to show that

$$h = \inf_{r,K} \{g_{r,K} : h(y) < r \text{ for all } y \in K\} = \vee \{g_{r,K} : h(y) < r \text{ for all } y \in K\}.$$

Thus, by definition of the "way below" relation, $g \ll f$ implies that

$$g \ge \vee\{g_{r_i,K_i} : h(y) < r_i \text{ for all } y \in K_i, i = 1, 2, \ldots, n\} =$$

$$\inf\{g_{r_i,K_i} : h(y) < r_i \text{ for all } y \in K_i, i = 1, 2, \ldots, n\},$$

and hence, $h \in \bigcap_{i=1}^{n} \{f : f(y) < r_i \text{ for all } y \in K_i\}$.

Let us show that this intersection is indeed a subset of A. Observe that for every $f \in USC(X)$, we have $g_{r_i,K_i} \ll f$ if $f(y) < r_i$ for all $y \in K_i$. Now let $\hat{f}$ be an arbitrary function from the intersection, i.e.,

$$\hat{f} \in \bigcap_{i=1}^{n} \{f : f(y) < r_i \text{ for all } y \in K_i\}.$$

Then,

$$g \stackrel{\text{def}}{=} \inf \{g_{r_i,K_i} : h(y) < r_i \text{ for all } y \in K_i, i = 1, \ldots, n\} \ll \hat{f}$$

and hence, by the representation of A,

$$\bigcap_{i=1}^{n} \{f : f(y) < r_i \text{ for all } y \in K_i, i = 1, \ldots, n\} \subseteq A.$$

To complete the proof, it is therefore sufficient to prove the following lemma:

Lemma 4. *For every $r \in (0, 1]$ and $K \in \mathcal{K}(X)$, we have*

$$\{f \in USC(X) : f(y) < r \text{ for all } y \in K\} = \bigcup_{\overset{\circ}{K_i} \supseteq K} \{f \in USC(X) : g_{r,K_i} \ll f\}.$$

Proof. Let $f \in USC(X)$ be such that $f(y) < r$ for all $y \in K$. Since f is usc, the set $A = \{x \in X : f(x) < r\}$ is an open set of X which contains K. Thus, there exists separating sets: an open set U and a compact set V such that $K \subseteq U \subseteq V \subseteq A$. Since $U \subseteq V$ and U is an open set, we thus conclude that $U \subseteq \overset{\circ}{V}$.

By definition of the function $g_{r,K}$, from $V \subseteq A$, we conclude that $g_{r,V} \ll f$. So,

$$\{f \in USC(X) : f(y) < r \text{ for all } y \in K\} \subseteq \bigcup_{\overset{\circ}{K_i} \supseteq K} \{f \in USC(X) : g_{r,K_i} \ll f\}.$$

Conversely, if $g_{r,K_i} \ll f$, where $\overset{\circ}{K} \supseteq K$, then for every $y \in K$, there exist r_y and K_y such that $y \in \overset{\circ}{K}_y$ and $f(z) < r_y \leq g_{r,K_i}(z)$ for all $z \in K_y$. In particular, for $z = y$, we $f(y) < r_y \leq g_{r,K_i}(y) = r$, so that $f(y) < r$ for all $y \in K$. The lemma is proven, and so is the theorem.

METRICS AND CHOQUET THEOREM FOR RANDOM FUZZY SETS

Resulting formal definition of a random fuzzy set. As we have mentioned, in general, a random object on a probability space $(\Omega, \mathcal{A}, P)$ is defined a mapping $x : \Omega \to O$ which is $\mathcal{A}$-$\mathcal{B}$-measurable with respect to an appropriate σ-field $\mathcal{B}$ of subsets of the set O of objects.

For closed fuzzy sets, the set $\mathcal{O}$ is the set $USC(X)$ of all semicontinuous functions, and the appropriate σ-algebra is the algebra $\mathcal{L}(X)$ of all Borel sets in Lawson topology.

Thus, we can define a *random fuzzy* (closed) *set* S on a probability space $(\Omega, \mathcal{A}, P)$ as a map $S : \Omega \to USC(X)$ which is $\mathcal{A}$-$\mathcal{L}(X)$-measurable.

Properties of the corresponding Lawson topology. From the general theory of continuous lattices, we can conclude that the space $USC(X)$ os a compact and Hausdorff topological space. When X is a LCHS space, then, similarly to the case of the set of all (crisp) closed sets $\mathcal{F}(X)$, we can prove that the set of all fuzzy closed sets $USC(X)$ is also second countable (see the proof below) and thus, metrizable.

Later in this section, we will discuss different metrics on $USC(X)$ which are compatible with this topology, and the corresponding Choquet theorem.

Theorem 5. *The topological space $USC(X)$ with the Lawson topology is second countable.*

Proof. It is known that for every continuous lattice, the Lawson topology is second countable if and only if the Scott topology has a countable base (Gierz et al., 1980). Thus, to prove our result, it is sufficient to prove that the Scott topology has a countable base.

In view of the results described in the previous section, the Scott topology has a base consisting of sets of the form

$$\bigcap_{i=1}^{n}\{f \in USC(X) : f(y) < r_i \text{ for all } y \in K_i\},$$

where $r_i \in (0, 1]$, $K_i \in \mathcal{K}(X)$, and $n \geq 0$. Let us denote

$$U(r_i, K_i) \stackrel{\text{def}}{=} \{f : f(y) < r_i \text{ for all } y \in K_i\}.$$

In these terms, the base of the Scott topology consists of the finite intersections $\bigcap\limits_{i=1}^{n} U(r_i, K_i)$.

Recall that since X is LCHS, X is normal, and there is a countable base $\mathcal{B}$ of the topological space $(X, \mathcal{O})$ such that for every $B \in \mathcal{B}$, the closure $\overline{B}$ is compact, and for any open set $G \in \mathcal{O}$, we have $G = \bigcup\limits_{j\in J} B_j$ for some $B_j \in \mathcal{B}$ with $\overline{B}_j \subseteq G$.

Our claim is that a countable base of $USC(X)$ consists of sets of the form $\bigcap\limits_{i=1}^{n} U\left(q_i, \bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}\right)$, where q_i are rational numbers from the interval $(0, 1]$, and $B_{ij} \in \mathcal{B}$.

It suffices to show that every neighborhood $\bigcap\limits_{i=1}^{n} U(r_i, K_i)$ of a closed fuzzy set $f \in USC(X)$ contains a sub-neighborhood $\bigcap\limits_{i=1}^{n} U\left(q_i, \bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}\right)$ (which still contains f).

Indeed, by definition of the sets $U(r_i, K_i)$, the fact that $f \in \bigcap\limits_{i=1}^{n} U(r_i, K_i)$ means that for every i and for every $y \in K_i$, we have $f(y) < r_i$. Let us denote by A_i the set of all the

values $x \in X$ for which $f(x) \geq r_i$: $A_i = \{x \in X : f(x) \geq r_i\}$. Since the function f is upper semicontinuous, the set A_i is closed. The sets A_i and K_i are both closed and clearly $A_i \cap K_i = \emptyset$. Since the space X is normal, there exists an open set G_i that separates A_i and K_i, i.e., for which $K_i \subseteq G_i$ and $A_i \cap G_i = \emptyset$. Due to the property of the base, we have $G_i = \bigcup\limits_{j \in J} B_{ij}$ with $\overline{B}_{ij} \subseteq G_i$. Thus, $K_i \subseteq \bigcup\limits_{j \in J} B_{ij}$. Since the set K_i is compact, from this open cover, we can extract a finite sub-cover, i.e., conclude that

$$K_i \subseteq \bigcup_{j=1}^{m_i} B_{ij} \subseteq \bigcup_{j=1}^{m_i} \overline{B}_{ij} \subseteq G_i.$$

From $A_i \cap G_i = \emptyset$, we can now conclude that $A_i \cap \left(\bigcup\limits_{j=1}^{m_i} \overline{B}_{ij} \right) = \emptyset$, implying that for every $y \in \bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}$, we have $y \notin A_i$, i.e., $f(y) < r_i$. Thus, $f(y) < r_i$ for any $y \in \bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}$.

Since f is usc, it attains its maximum on $\bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}$ at some point $y_{\max}$. For this maximum value, we therefore also have $f(y_{\max}) < r_i$ and therefore, there exists a rational number q_i such that $f(y_{\max}) < q_i < r_i$. Since the value $f(y_{\max})$ is the maximum, we conclude that $f(y) \leq f(y_{\max})$ for all other $y \in \bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}$. Thus, for all such y, we have $f(y) < q_i$. This means that $f \in \bigcap\limits_{i=1}^{n} U\left(q_i, \bigcup\limits_{j=1}^{m_i} \overline{B}_{ij}\right)$.

It is easy to show that

$$\bigcap_{i=1}^{n} U\left(q_i, \bigcup_{j=1}^{m_i} \overline{B}_{ij}\right) \subseteq \bigcap_{i=1}^{n} U(r_i, K_i).$$

The theorem is proven.

Towards defining metrics on $USC(X)$. We have proven that the Lawson topology on $USC(X)$ is metrizable, i.e., that there exists a metric with is compatible with this topology. From the viewpoint of possible applications, it is desirable to give an explicit description of such a metric.

In Section 4, we used the point compactification procedure to explicit describe a specific metric compatible with the Lawson topology on the set $\mathcal{F}(X)$ of all closed (crisp) sets. Let us show that for the class $USC(X)$ of closed fuzzy sets, we can define a similar metric if we identify these closed fuzzy sets with their *hypographs*, i.e., informally, areas below their graphs.

Formally, for every function $f : X \to [0, 1]$, its *hypograph* $Hyp(f)$ is defined as

$$Hyp(f) = \{(x, \alpha) \in X \times [0, 1] : f(x) \geq \alpha\}.$$

Every hypograph is a subset of $X \times [0, 1]$. Since we consider usc functions, each hypograph is closed.

Let $HYP(X)$ denote the set of all hypographs of all functions $f \in USC(X)$. Then, $f \to Hyp(f)$ is a bijection from $USC(X)$ to $HYP(X)$; see (Nguyen, Wang, and Wei, in press). One can easily check that the set $HYP(X)$ is a closed subset of $\mathcal{F}(X \times [0,1])$. Note that since X is a LCHS space, the set $\mathcal{F}(X \times [0,1])$ is also a LCHS space, so the set $\mathcal{F}(X \times [0,1])$ of all its closed subsets is a topological space with a canonical Lawson topology. Thus, according to Section 4, $\mathcal{F}(X \times [0,1])$ has a compatible metric H. Then the induced metric on $HYP(X)$ is also compatible with the induced Lawson topology (or hit-or-miss topology) on $HYP(X)$. The only things that remains to be proven is that $(HYP(X), H)$ is homeomorphic to $(USC(X), \mathcal{L})$, where $\mathcal{L}$ denotes the Lawson topology on $USC(X)$. This result was proven in (Nguyen and Tran, 2007).

Finally, a Choquet theorem for $USC(X)$ can be obtained by embedding $USC(X)$ into $\mathcal{F}(X \times [0,1])$ via hypographs and using Choquet theorem for random closed sets on $X \times [0,1]$. For more details, see (Nguyen, Wang, and Wei, in press).

TOWARDS PRACTICAL APPLICATIONS

From the general theory to computationally feasible practical applications. In the previous sections, we described a general theory of random fuzzy sets, a theory motivated by (and tailored towards) potential applications. Our main motivation is to prepare the background for as wide a range of applications as possible. Because of this objective, we tried our best to make our theory as general as possible.

Because of this same application objective, we also tried our best to make the resulting techniques as computation-friendly as possible. However, as we will see in this section, the resulting problems are computationally difficult even in the simplest cases. Because of this difficulty, in this chapter, we will mainly concentrate on such simple cases.

Practical need for random sets and random fuzzy sets: reminder. We have started this paper with explaining the practical motivation for random sets and random fuzzy sets.

This motivation is that due to measurement uncertainty (or uncertainty of expert estimates), often, instead of the actual values x_i of the quantities of interest, we only know the intervals $\mathbf{x_i} = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$, where $\widetilde{x}_i$ is the known approximate value and Δ_i is the upper bound on the approximation error (provided, for measurements, by the manufacturer of the measuring instrument).

These intervals can be viewed as *random intervals*, i.e., as samples from the interval-valued random variable. In such situations, instead of the exact value of the sample statistics such as covariance $C[x,y]$, we can only have an interval $\mathbf{C}[x,y]$ of possible values of this statistic.

The need for such random intervals is well recognized, and there has already been a lot of related research; see, e.g., (Möller and Beer, 2004). In this approach, the uncertainty in a vector quantity $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ is usually described by describing intervals of possible values of each of its components. This is equivalent to describing the set of all possible values of x as a *box* ("multi-dimensional interval") $[\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_n, \overline{x}_n]$. However, the resulting data processing problem are often very challenging, and there is still a large room for further development.

One such need comes from the fact that uncertainty is often much more complex than intervals. For example, for the case of several variables, instead of an multi-dimensional interval, we may have a more complex set $S \subseteq \mathbb{R}^d$. In such a situation, we need a more general theory of random sets.

We have also mentioned that to get a more adequate description of expert estimates, we need to supplement the set S of possible values of the quantity (or quantities) of interest with describing the sets S_α which contain values which are possible with a certain degree α. In such situations, we need to consider random fuzzy sets.

What is needed for practical applications: an outline of this section. As we have just recalled, there is a practical need to consider random sets and random fuzzy sets. In order to apply the corresponding theory, we first need to estimate the actual distribution of random sets or random fuzzy sets from the observations.

In other words, we need to develop *statistical* techniques for random sets and random fuzzy sets. In this section, we start with a reminder about traditional number-valued and vector-valued statistical techniques, and the need for extending these techniques to random sets and random fuzzy sets. Then, we overview the main sources of the corresponding data uncertainty and techniques for dealing with the corresponding uncertainty. This will prepare us for the case study described in the following section.

Traditional statistics: brief reminder. In traditional statistics, we assume that the observed values are independent identically distributed (i.i.d.) variables $x_1, \ldots, x_n, \ldots$, and we want to find statistics $C_n(x_1, \ldots, x_n)$ that would approximate the desired parameter C of the corresponding probability distribution.

For example, if we want to estimate the mean E, we can take the arithmetic average $E_n[x] = \dfrac{x_1 + \ldots + x_n}{n}$. It is known that as $n \to \infty$, this statistic tends (with probability 1) to the desired mean: $E_n[x] \to E$. Similarly, the sample variance $V_n[x] = \dfrac{1}{n-1} \cdot \sum\limits_{i=1}^{n} (x_i - E_n[x])^2$ tends to the actual variance V, the sample covariance

$$C_n[x, y] = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - E_n[x]) \cdot (y_i - E_n[y])$$

between two different samples tends to the actual covariance C, etc.

Coarsening: a source of random sets. In traditional statistics, we implicitly assume that the values x_i are directly observable. In real life, due to (inevitable) measurement uncertainty, often, what we actually observe is a *set* S_i that contains the actual (unknown) value of x_i. This phenomenon is called *coarsening*; see, e.g., (Heitjan and Rubin, 1991). Due to coarsening, instead of the actual values x_i, all we know is the sets $X_1, \ldots, X_n, \ldots$ that are known the contain the actual (un-observable) values x_i: $x_i \in X_i$.

Statistics based on coarsening. The sets $X_1, \ldots, X_n, \ldots$ are i.i.d. *random sets*. We want to find statistics of these random sets that would enable us to approximate the desired

parameters of the original distribution x. Here, a statistic $S_n(X_1, \ldots, X_n)$ transform n sets $X_1, \ldots, X_n$ into a new set. We want this statistic $S_n(X_1, \ldots, X_n)$ to tend to a limit set L as $n \to \infty$, and we want this limit set L to contain the value of the desired parameter of the original distribution.

For example, if we are interested in the mean $E[x]$, then we can take $S_n = (X_1 + \ldots + X_n)/n$ (where the sum is the Minkowski – element-wise – sum of the sets). It is possible to show that, under reasonable assumptions, this statistic tends to a limit L, and that $E[x] \in L$. This limit can be viewed, therefore, as a set-based average of the sets $X_1, \ldots, X_n$.

Important issue: computational complexity. There has been a lot of interesting theoretical research on set-valued random variables and corresponding statistics. In many cases, the corresponding statistics have been designed, and their asymptotical properties have been proven; see, e.g., (Goutsias, Mahler, and Nguyen, 1997), (Li, Ogura, and Kreinovich, 2002), (Nguyen, 2006) and references therein.

In many such situations, the main obstacle to a practical use of these statistics is that going from random numbers to random sets drastically increases the computational complexity – hence, the running time – of the required computations. It is therefore desirable to come up with new, faster algorithms for computing such set-values heuristics.

Sources of uncertainty: general reminder. Traditional engineering statistical formulas assume that we know the *exact* values x_i of the corresponding quantity. In practice, these values come either from measurements or from expert estimates. In both case, we get only *approximations* $\widetilde{x}_i$ to the actual (unknown) values x_i.

When we use these approximate values $\widetilde{x}_i \neq x_i$ to compute the desired statistical characteristics such as E and V, we only get approximate valued $\widetilde{E}$ and $\widetilde{V}$ for these characteristics. It is desirable to estimate the accuracy of these approximations.

Case of measurement uncertainty. Measurements are never 100% accurate. As a result, the result $\widetilde{x}$ of the measurement is, in general, different from the (unknown) actual value x of the desired quantity. The difference $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ between the measured and the actual values is usually called a *measurement error*.

The manufacturers of a measuring device usually provide us with an upper bound Δ for the (absolute value of) possible errors, i.e., with a bound Δ for which we guarantee that $|\Delta x| \leq \Delta$. The need for such a bound comes from the very nature of a measurement process: if no such bound is provided, this means that the difference between the (unknown) actual value x and the observed value $\widetilde{x}$ can be as large as possible.

Since the (absolute value of the) measurement error $\Delta x = \tilde{x} - x$ is bounded by the given bound Δ, we can therefore guarantee that the actual (unknown) value of the desired quantity belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

Traditional probabilistic approach to describing measurement uncertainty. In many practical situations, we not only know the interval $[-\Delta, \Delta]$ of possible values of the measurement error; we also know the probability of different values Δx within this interval;

see, e.g., (Rabinovich, 2005).

In practice, we can determine the desired probabilities of different values of Δx by comparing the results of measuring with this instrument with the results of measuring the same quantity by a standard (much more accurate) measuring instrument. Since the standard measuring instrument is much more accurate than the one use, the difference between these two measurement results is practically equal to the measurement error; thus, the empirical distribution of this difference is close to the desired probability distribution for measurement error.

Interval approach to measurement uncertainty. As we have mentioned, in many practical situations, we do know the probabilities of different values of the measurement error. There are two cases, however, when this determination is not done:

- First is the case of cutting-edge measurements, e.g., measurements in fundamental science. When a Hubble telescope detects the light from a distant galaxy, there is no "standard" (much more accurate) telescope floating nearby that we can use to calibrate the Hubble: the Hubble telescope is the best we have.
- The second case is the case of measurements on the shop floor. In this case, in principle, every sensor can be thoroughly calibrated, but sensor calibration is so costly – usually costing ten times more than the sensor itself – that manufacturers rarely do it.

In both cases, we have no information about the probabilities of Δx; the only information we have is the upper bound on the measurement error.

In this case, after performing a measurement and getting a measurement result $\widetilde{x}$, the only information that we have about the actual value x of the measured quantity is that it belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$. In this situation, for each i, we know the interval $\mathbf{x}_i$ of possible values of x_i, and we need to find the ranges $\mathbf{E}$ and $\mathbf{V}$ of the characteristics E and V over all possible tuples $x_i \in \mathbf{x}_i$.

Case of expert uncertainty. An expert usually describes his/her uncertainty by using words from the natural language, like "most probably, the value of the quantity is between 6 and 7, but it is somewhat possible to have values between 5 and 8". To formalize this knowledge, it is natural to use *fuzzy set theory*, a formalism specifically designed for describing this type of informal ("fuzzy") knowledge (Klir and Yuan, 1995), (Nguyen and Walker, 2006).

As a result, for every value x_i, we have a fuzzy set $\mu_i(x_i)$ which describes the expert's prior knowledge about x_i: the number $\mu_i(x_i)$ describes the expert's degree of certainty that x_i is a possible value of the i-th quantity.

As we have mentioned earlier, an alternative user-friendly way to represent a fuzzy set is by using its α-cuts $\{x_i : \mu_i(x_i) \geq \alpha\}$. In these terms, a fuzzy set can be viewed as a nested family of intervals $[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]$ corresponding to different level α.

Estimating statistics under fuzzy uncertainty: precise formulation of the problem. In general, we have fuzzy knowledge $\mu_i(x_i)$ about each value x_i; we want to find

the fuzzy set corresponding to a given characteristic $y = C(x_1, \ldots, x_n)$. Intuitively, the value y is a reasonable value of the characteristic if $y = C(x_1, \ldots, x_n)$ for some reasonable values x_i, i.e., if for some values $x_1, \ldots, x_n$, x_1 is reasonable, and x_2 is reasonable, ..., and $y = C(x_1 \ldots, x_n)$. If we interpret "and" as min and "for some" ("or") as max, then we conclude that the corresponding degree of certainty $\mu(y)$ in y is equal to $\mu(y) = \max\{\min(\mu_1(x_1), \ldots, \mu_n(x_n)) : C(x_1, \ldots, x_n) = y\}$.

Reduction to the case of interval uncertainty. It is known that the above formula (called *extension principle*) can be reformulated as follows: for each α, the α-cut $\mathbf{y}(\alpha)$ of y is equal to the range of possible values of $C(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i(\alpha)$ for all i. Thus, from the computational viewpoint, the problem of computing the statistical characteristic under fuzzy uncertainty can be reduced to the problem of computing this characteristic under interval uncertainty; see, e.g., (Dubois, Fargier, and Fortin, 2005).

In view of this reduction, in the following text, we will consider the case of interval (and set) uncertainty.

Estimating statistics under interval uncertainty: a problem. In the case of interval uncertainty, instead of the true values $x_1, \ldots, x_n$, we only know the intervals $\mathbf{x}_1 = [\underline{x}_1, \overline{x}_1], \ldots, \mathbf{x}_n = [\underline{x}_n, \overline{x}_n]$ that contain the (unknown) true values of the measured quantities. For different values $x_i \in \mathbf{x}_i$, we get, in general, different values of the corresponding statistical characteristic $C(x_1, \ldots, x_n)$. Since all values $x_i \in \mathbf{x}_i$ are possible, we conclude that all the values $C(x_1, \ldots, x_n)$ corresponding to $x_i \in \mathbf{x}_i$ are possible estimates for the corresponding statistical characteristic. Therefore, for the interval data $\mathbf{x}_1, \ldots, \mathbf{x}_n$, a reasonable estimate for the corresponding statistical characteristic is the range

$$C(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{C(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

We must therefore modify the existing statistical algorithms so that they compute, or bound these ranges.

Estimating mean under interval uncertainty. The arithmetic average E is a monotonically increasing function of each of its n variables $x_1, \ldots, x_n$, so its smallest possible value $\underline{E}$ is attained when each value x_i is the smallest possible ($x_i = \underline{x}_i$) and its largest possible value is attained when $x_i = \overline{x}_i$ for all i. In other words, the range $\mathbf{E}$ of E is equal to $[E(\underline{x}_1, \ldots, x_n), E(\overline{x}_1, \ldots, \overline{x}_n)]$. In other words, $\underline{E} = \frac{1}{n} \cdot (\underline{x}_1 + \ldots + \underline{x}_n)$ and $\overline{E} = \frac{1}{n} \cdot (\overline{x}_1 + \ldots + \overline{x}_n)$.

Estimating variance under interval uncertainty. It is known that the problem of computing the exact range $\mathbf{V} = [\underline{V}, \overline{V}]$ for the variance V over interval data $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ is, in general, NP-hard; see, e.g., (Kreinovich et al., 2006), (Kreinovich et al., 2007). Specifically, there is a polynomial-time algorithm for computing $\underline{V}$, but computing $\overline{V}$ is, in general, NP-hard.

Comment. NP-hard means, crudely speaking, that no feasible algorithm can compute the exact value of $\overline{V}$ for all possible intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$; see, e.g., (Kreinovich et al., 1997).

In many practical situations, there are efficient algorithms for computing $\overline{V}$: e.g., an $O(n \cdot \log(n))$ time algorithm exists when no two narrowed intervals $[x_i^-, x_i^+]$, where $x_i^- \stackrel{\text{def}}{=} \widetilde{x}_i - \dfrac{\Delta_i}{n}$

and $x_i^+ \stackrel{\text{def}}{=} \widetilde{x}_i + \dfrac{\Delta_i}{n}$, are proper subsets of one another, i.e., when $[x_i^-, x_i^+] \not\subseteq (x_j^-, x_j^+)$ for all i and j (Dantsin et al., 2006).

What can be done if we cannot effectively compute the exact range. As we have just mentioned, the problem of computing statistical characteristics under interval uncertainty is often NP-hard – which means, crudely speaking, that we cannot efficiently compute the exact range for these characteristics.

A natural solution is as follows: since we cannot compute the *exact* range, we should try to find an *enclosure* for this range. Computing the range $C(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of a function $C(x_1, \ldots, x_n)$ based on the input intervals $\mathbf{x}_i$ is called *interval computations*; see, e.g., (Jaulin et al., 2001).

Interval computations techniques: brief reminder. Historically the first method for computing the enclosure for the range is the method which is sometimes called "straightforward" interval computations. This method is based on the fact that inside the computer, every algorithm consists of elementary operations (arithmetic operations, min, max, etc.). For each elementary operation $f(a, b)$, if we know the intervals $\mathbf{a}$ and $\mathbf{b}$ for a and b, we can compute the exact range $f(\mathbf{a}, \mathbf{b})$. The corresponding formulas form the so-called *interval arithmetic*. For example,

$$[\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}]; \quad [\underline{a}, \overline{a}] - [\underline{b}, \overline{b}] = [\underline{a} - \overline{b}, \overline{a} - \underline{b}];$$

$$[\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] = [\min(\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b}), \max(\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b})].$$

In straightforward interval computations, we repeat the computations forming the program C for computing $C(x_1, \ldots, x_n)$ step-by-step, replacing each operation with real numbers by the corresponding operation of interval arithmetic. It is known that, as a result, we get an enclosure $\mathbf{Y} \supseteq C(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ for the desired range.

In some cases, this enclosure is exact. In more complex cases (see examples below), the enclosure has excess width.

There exist more sophisticated techniques for producing a narrower enclosure, e.g., a centered form method. However, for each of these techniques, there are cases when we get an excess width. (Reason: as have mentioned, the problem of computing the exact range is known to be NP-hard even for population variance.)

CASE STUDY: A BIOINFORMATICS PROBLEM

In this section, we describe an example of a practical applications. This example was first outlined in (Kreinovich et al., 2007) and (Xiang, 2007). For other applications, see (Kreinovich et al., 2006), (Kreinovich et al., 2007) and references therein.

Description of the case study. In cancer research, it is important to find out the genetic difference between the cancer cells and the healthy cells. In the ideal world, we should be able to have a sample of cancer cells, and a sample of healthy cells, and thus directly measure the concentrations c and h of a given gene in cancer and in healthy cells. In reality, it is very difficult to separate the cells, so we have to deal with samples that contain both cancer

and normal cells. Let y_i denote the result of measuring the concentration of the gene in i-th sample, and let x_i denote the percentage of cancer cells in i-th sample. Then, we should have $x_i \cdot c + (1 - x_i) \cdot h \approx y_i$ (approximately equal because there are measurement errors in measuring y_i).

Let us first consider an idealized case in which we know the exact percentages x_i. In this case, we can find the desired values c and h by solving a system of linear equations $x_i \cdot c + (1 - x_i) \cdot h \approx y_i$ with two unknowns c and h.

It is worth mentioning that this system can be somewhat simplified if instead of c, we consider a new variable $a \stackrel{\text{def}}{=} c - h$. In terms of the new unknowns a and h, the system takes the following form: $a \cdot x_i + h \approx y_i$.

The errors of measuring y_i are normally i.i.d. random variables, so to estimate a and h, we can use the Least Squares Method (LSM) $\sum\limits_{i=1}^{n} (a \cdot x_i + h - y_i)^2 \to \min\limits_{a,h}$. According to LSM, we have $a = \dfrac{C[x, y]}{V[x]}$ and $h = E[y] - a \cdot E[x]$, where $E[x] = \dfrac{1}{n}(x_1 + \ldots + x_n)$, $E[y] = \dfrac{1}{n}(y_1 + \ldots + y_n)$, $V[x] = \dfrac{1}{n-1} \cdot \sum\limits_{i=1}^{n} (x_i - E[x])^2$, and $C[x, y] = \dfrac{1}{n-1} \cdot \sum\limits_{i=1}^{n} (x_i - E[x]) \cdot (y_i - E[y])$. Once we know $a = c - h$ and h, we can then estimate c as $a + h$.

The problem is that the concentrations x_i come from experts who manually count different cells, and experts can only provide interval bounds on the values x_i such as $x_i \in [0.7, 0.8]$ (or even only fuzzy bounds). Different values of x_i in the corresponding intervals lead to different values of a and h. It is therefore desirable to find the range of a and h corresponding to all possible values $x_i \in [\underline{x}_i, \overline{x}_i]$.

Comment. Our motivation for solving this problem comes from bioinformatics, but similar problems appear in various practical situations where measurements with uncertainties are available and statistical data is to be processed.

Linear approximation. Let $\widetilde{x}_i = (\underline{x}_i + \overline{x}_i)/2$ be the midpoint of i-th intervals, and let $\Delta_i = (\overline{x}_i - \underline{x}_i)/2$ be its half-width. For a, we have

$$\frac{\partial a}{\partial x_i} = \frac{1}{(n-1) \cdot V[x]} \cdot (y_i - E[y] - 2a \cdot x_i + 2a \cdot E[x]).$$

We can use the formula $E[y] = a \cdot E[x] + h$ to simplify this expression, resulting in $\Delta_a = \dfrac{1}{(n-1) \cdot V[x]} \sum\limits_{i=1}^{n} |\Delta y_i - a \cdot \Delta x_i| \cdot \Delta_i$, where we denoted $\Delta y_i \stackrel{\text{def}}{=} y_i - a \cdot x_i - h$ and $\Delta x_i \stackrel{\text{def}}{=} x_i - E[x]$.

Since $h = E[y] - a \cdot E[x]$, we have $\dfrac{\partial h}{\partial x_i} = -\dfrac{\partial a}{\partial x_i} \cdot E[x] - \dfrac{1}{n} \cdot a$, so $\Delta_h = \sum\limits_{i=1}^{n} \left| \dfrac{\partial h}{\partial x_i} \right| \cdot \Delta_i$.

Prior estimation of the resulting accuracy. The above formulas provide us with the accuracy *after* the data has been processed. It is often desirable to have an estimate *prior* to measurements, to make sure that we will get c and h with desired accuracy.

The difference Δy_i is a measurement error, so it is normally distributed with 0 mean and

standard deviation $\sigma(y)$ corresponding to the accuracy of measuring y_i. The difference Δx_i is distributed with 0 mean and standard deviation $\sqrt{V[x]}$. For estimation purposes, it is reasonable to assume that the values Δx_i are also normally distributed. It is also reasonable to assume that the errors in x_i and y_i are uncorrelated, so the linear combination $\Delta y_i - a \cdot \Delta x_i$ is also normally distributed, with 0 mean and variance $\sigma_y^2 + a^2 \cdot V[x]$. It is also reasonable to assume that all the values Δ_i are approximately the same: $\Delta_i \approx \Delta$.

For normal distribution ξ with 0 mean and standard deviation σ, the mean value of $|\xi|$ is equal to $\sqrt{2/\pi} \cdot \sigma$. Thus, the absolute value $|\Delta y_i - a \cdot \Delta x_i|$ of the above combination has a mean value $\sqrt{2/\pi} \cdot \sqrt{\sigma_y^2 + a^2 \cdot V[x]}$. Hence, the expected value of Δ_a is equal to $\dfrac{2}{\pi} \cdot \dfrac{\sqrt{\sigma_y^2 + a^2 \cdot V[x]} \cdot \Delta}{V[x]}$.

Since measurements are usually more accurate than expert estimates, we have $\sigma_y^2 \ll V[x]$, hence $\Delta_a \approx \dfrac{2}{\pi} \cdot a \cdot \Delta$.

Similar estimates can be given for Δ_h.

Why not get exact estimates? Because in general, finding the exact range is NP-hard. Let us show that in general, finding the exact range for the ratio $C[x, y]/V[x]$ is an NP-hard problem; this proof was first presented in (Kreinovich et al., 2007).

The proof is similar to the proof that computing the range for the variance is NP-hard (Ferson et al., 2005): namely, we reduce a partition problem (known to be NP-hard) to our problem. In the partition problem, we are given m positive integers $s_1, \ldots, s_m$, and we must check whether there exist values $\varepsilon_i \in \{-1, 1\}$ for which $\sum\limits_{i=1}^{m} \varepsilon_i \cdot s_i = 0$. We will reduce this problem to the following problem: $n = m + 2$, $y_1 = \ldots = y_m = 0$, $y_{m+1} = 1$, $y_{m+2} = -1$, $x_i = [-s_i, s_i]$ for $i \le m$, $x_{m+1} = 1$, and $x_{m+2} = -1$. In this case, $E[y] = 0$, so $C[x, y] = \dfrac{1}{n-1} \sum\limits_{i=1}^{n} x_i \cdot y_i - \dfrac{n}{n-1} \cdot E[x] \cdot E[y] = \dfrac{2}{m+2}$. Therefore, $C[x, y]/V[x] \to \min$ if and only if $V[x] \to \max$.

Here, $V[x] = \dfrac{1}{m+1} \cdot \left(\sum\limits_{i=1}^{m} x_i^2 + 2 \right) - \dfrac{m+2}{m+1} \cdot \left(\dfrac{1}{m+2} \cdot \sum\limits_{i=1}^{m} x_i \right)^2$. Since $|x_i| \le s_i$, we always have $V[x] \le V_0 \stackrel{\text{def}}{=} \dfrac{1}{m+1} \cdot \left(\sum\limits_{i=1}^{m} s_i^2 + 2 \right)$, and the only possibility to have $V[x] = V_0$ is when $x_i = \pm s_i$ for all i and $\sum x_i = 0$. Thus, $V[x] = V_0$ if and only if the original partition problem has a solution. Hence, $C[x, y]/V[x] = \dfrac{2}{\sum s_i^2 + 2}$ if and only if the original instance of the partition problem has a solution.

The reduction is proven, so our problem is indeed NP-hard.

Comment 1. In this proof, we consider the case when the values x_i can be negative and

larger than 1, while in bioinformatics, x_i is always between 0 and 1. However, we can easily modify this proof: First, we can shift all the values x_i by the same constant to make them positive; shift does not change neither $C[x,y]$ nor $V[x]$. Second, to make the positive values ≤ 1, we can then re-scale the values x_i ($x_i \to \lambda \cdot x_i$), thus multiplying $C[x,y]/V[x]$ by a known constant.

As a result, we get new values $x_i' = \frac{1}{2} \cdot (1 + x_i/K)$, where $K \stackrel{\text{def}}{=} \max s_i$, for which $x_i' \in [0,1]$ and the problem of computing $C[x,y]/V[x]$ is still NP-hard.

Comment 2. Since we cannot compute the exact range, what can we do to compute the more accurate estimates for the range than those provided by linear approximation? One possibility is to use known algorithms to find the ranges for $C[x,y]$ and for $V[x]$ (Kreinovich et al., 2006), (Kreinovich et al., 2007), and then use the division operation from interval arithmetic to get the interval that is guaranteed to contain $C[x,y]/V[x]$.

Acknowledgments. This work was supported in part by NSF grants HRD-0734825 and EAR-0225670, by Texas Department of Transportation grant No. 0-5453, and by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08. The authors are thankful to the anonymous referees for their valuable suggestions.

REFERENCES

Dantsin, E., Kreinovich, V., Wolpert, A., & Xiang, G. (2006) Population Variance under Interval Uncertainty: A New Algorithm, *Reliable Computing*, 12(4), 273-280.

Dubois, D., Fargier, H., & Fortin, J. (2005) The empirical variance of a set of fuzzy intervals, *Proceedings of the 2005 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2005*, Reno, Nevada, May 22–25, pp. 885-890.

Fell, J. M. G. (1961) The structure of algebras of operator fields, *Acta Math.*, 101:19-38.

Fell, J. M. G. (1962) A Hausdorff topology for the closed sets of a locally compact non-Hausdorff space, *Proc. Amer. Math. Soc.*, 13:472–476.

Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., & Aviles, M. (2005) Exact Bounds on Finite Populations of Interval Data, *Reliable Computing*, 11(3): 207-233.

Gierz, G., Hofmann, K. H., Keimel, K., Lawson, J. D., Mislove, M., & Scott, D. S. (1980) *A Compendium of Continuous Lattices*, Springer-Verlag, Berlin, Heidelberg, New York.

Goutsias, J., Mahler, R. P. S., & Nguyen, H. T. eds. (1997) *Random Sets: Theory and Applications*, Springer-Verlag, New York.

Heitjan D. F., & Rubin, D. B. (1991) Ignorability and coarse data, *Ann. Stat.*, 19(4):2244-2253.

Jaulin L., Kieffer, M., Didrit, O., & Walter, E. (2001) *Applied Interval Analysis*, Springer-Verlag, London.

Klir, G., & Yuan, B. (1995) *Fuzzy sets and fuzzy logic: theory and applications*, Prentice

Hall, Upper Saddle River, New Jersey.

Kreinovich V., Lakeyev, A., Rohn, J., & Kahl, P. (1997) *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht.

Kreinovich, V., Longpré, L., Starks, S. A., Xiang, G., Beck, J., Kandathi, R., Nayak, A., Ferson, S., & Hajagos, J. (2007) Interval Versions of Statistical Techniques, with Applications to Environmental Analysis, Bioinformatics, and Privacy in Statistical Databases", *Journal of Computational and Applied Mathematics*, 199(2):418-423.

Kreinovich, V., Xiang, G., Starks, S. A., Longpre, L., Ceberio, M., Araiza, R., Beck, J., Kandathi, R., Nayak, A., Torres, R., & Hajagos, J. (2006) Towards combining probabilistic and interval uncertainty in engineering calculations: algorithms for computing statistics under interval uncertainty, and their computational complexity, *Reliable Computing*, 12(6):471-501.

Li, S., Ogura, Y., & Kreinovich, V. (2002) *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer, Dordrecht.

Mathéron, G. (1975) *Random Sets and Integral Geometry*, J. Wiley.

Möller, & Beer, M. (2004) *Fuzzy randomness: Uncertainty in Civil Engineering and Computational Mechanics*, Springer-Verlag.

Nguyen, H. T. (2006) *An Introduction to Random Sets*, Chapman and Hall/CRC Press.

Nguyen, H. T., & Kreinovich, V. (1996) Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, In: Kearfott, R. B., & Kreinovich, V. (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, pp. 245-290.

Nguyen, H. T. & Tran, H. (2007) On a continuous lattice approach to modeling of coarse data in system analysis, *Journal of Uncertain Systems*, 1(1):62-73.

Nguyen, H. T., & Walker, E. A. (2006) *A First Course in Fuzzy Logic* (3rd edition), Chapman and Hall/CRC Press.

Nguyen, H. T., Wang, Y., & Wei, G. (in press) On Choquet theorem for random upper semicontinuous functions, *International Journal of Approximate Reasoning*.

Rabinovich, S. (2005) *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York.

Rockafellar, R. T., & West, J. B. (1984) Variational systems an introduction. In *Springer Lecture Notes in Mathematics*, Vol. 1091, 1–54.

Wei, G., & Wang, Y. (in press) On metrization of the hit-or-miss topology using Alexandroff compactification, *International Journal of Approximate Reasoning*.

Xiang, G. (2007) Interval uncertainty in bioinformatics, *Abstracts of the New Mexico Bioinformatics Symposium*, Santa Fe, Mexico, March 8–9, 2007, p. 25.