Information Entropy and Co–entropy of Crisp and Fuzzy granulations

D. Bianucci and G. Cattaneo and D. Ciucci

Università degli Studi di Milano–Bicocca
Dipartimento di Informatica, Sistemistica e Comunicazione,
Via Bicocca degli Arcimboldi 8, 20126 Milano (Italy)
{bianucci, cattang, ciucci}@disco.unimib.it

Abstract. The standard approach to information entropy applied to partitions of a universe is equivalently formulated as the entropy of the corresponding crisp identity resolutions, interpreted as crisp granulations, by the corresponding characteristic functionals. Moreover, in this crisp context the co–entropy notion is introduced.
The extension to the case of fuzzy identity resolutions, a particular case of fuzzy granulation, is studied.

1 Entropy of Abstract Discrete Probability Distributions

In this section we briefly discuss the *abstract* approach to information theory, involving suitable finite sequences of numbers from the real unit interval $[0, 1]$, each of which can be interpreted as a probability of occurrence of something, without any reference to a concrete universe X. To be precise, a length N *probability distribution* is a vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_N)$ in which: (pd-1)   every $p_1 \geq 0$ and (pd-2) $\sum_{i=1}^{n} p_i = 1$. In this abstract context, a length N *random variable* is a real vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_N)$ in which each component is a real number: $a_i \in \mathbb{R}$ for any i. For a fixed length N random variable $\boldsymbol{a}$ and a length N probability distribution $\boldsymbol{p}$, the numbers a_i are interpreted as *possible values* of $\boldsymbol{a}$ and the quantities p_i as the probability of occurrence of the event "$\boldsymbol{a} = a_i$" (thus, p_i can be considered as a simplified notation of $p(a_i)$) (see [1, p.5]).

Hence, the *average* (or *mean*) value of the random variable $\boldsymbol{a}$ with respect to a probability distribution $\boldsymbol{p}$ is given by $Av(\boldsymbol{a}, \boldsymbol{p}) = \sum_{i=1}^{N} a_i \cdot p_i$.

In particular, to any probability distribution $\boldsymbol{p} = (p_1, p_2, \ldots, p_N)$ it is possible to associate the *information random variable* $\boldsymbol{I}[\boldsymbol{p}] = (I(p_1), I(p_2), \ldots, I(p_N))$, where for every $p \in (0, 1]$ it is $I(p) := -\log(p)$, whose mean value with respect to the probability distribution $\boldsymbol{p}$, called the *entropy* of the probability distribution and denoted by $H(\boldsymbol{p})$, is explicitly expressed by the formula (with the convention $0 \log 0 = 0$):

$$H(\boldsymbol{p}) = -\sum_{i=1}^{N} p_i \log p_i \tag{1}$$

with $0 \leq H(\boldsymbol{p}) \leq \log N$. Let us recall that the real number $I(p) := -\log(p)$ is a measure (called the Hartley *measure* [2]) of the *uncertainty* due to the knowledge

of a probability since if the probability is 1, then there is no uncertainty and so its corresponding measure is 0. Moreover, any probability different from 1 (and 0) is linked to some uncertainty whose measure is greater than 0 in such a way that the lower is the probability, the greater is the corresponding uncertainty measure. Hence, the entropy of a probability distribution $\boldsymbol{p}$ can be considered as a quantity which in a reasonable way measures the *average amount of uncertainty* associated with this distribution, expressed as the mean value of the corresponding information random variable $\boldsymbol{I}[\boldsymbol{p}]$.

2 Entropy and Co–Entropy of Partitions

Now we apply the just discussed notion of *information entropy* to the concrete case of partitions of a *finite* nonempty universe X. A partition of X is a finite collection $\pi = \{A_1, A_2, \ldots, A_N\}$ of nonempty subsets A_i of X which are pairwise disjoints and whose set theoretic union is X. Elements A_i forming the partition π are considered as *granules* of some knowledge associated with the partition. As it is well known, a partition π is equivalently described by an equivalence (reflexive, symmetric and transitive) relation on X formally written as $(x, y)\mathcal{R}$ iff $\exists A_j \in \pi : x, y \in A_j$, and this equivalence relation expresses the fact that two objects x and y of the universe cannot be *distinguished* relatively to the knowledge supported by the partition. The equivalence relation is in this way interpreted as an *indistinguishability* relation (see [5], [6], [7]). In this sense a partition furnishes a *granulation* of the universe X by "crisp" granules, also if in the sequel we adopt a weaker notion of granulation linked to a measure distribution assigned to every equivalence class A_i of the partition.

Given a partition π, the subsets $A_i \in \pi$ are the *elementary events* of the *measure distribution* $\boldsymbol{m}(\pi) = (|A_1|, |A_2|, \ldots, |A_N|)$, where the measure of the event A_i is an application to it of the so–called *counting measure* $m_c(Y) := |Y|$ (the cardinality measure) for any arbitrary subset Y of X. This measure distribution satisfies the conditions: (md-1) every $|A_i| > 0$; (md-2) its *total measure* is $M(\pi) := \sum_{i=1}^{N} |A_i| = |X|$. The *probability of occurrence* of the event A_i is then given by $p(A_i) = \frac{|A_i|}{|X|}$. In this way we have generated the *probability distribution* $\boldsymbol{p}(\pi) := (p(A_1), p(A_2), \ldots, p(A_N))$, depending from the partition π. This probability distribution satisfies the conditions: (pd-1) every $p(A_i) > 0$; (pd-2) its *total probability* is $P(\pi) := \sum_{i=1}^{N} p(A_i) = 1$. According to (1), the entropy of this probability distribution, simply written as $H(\pi)$ instead of $H(\boldsymbol{p}(\pi))$, is then

$$H(\pi) = -\sum p(A_i) \log p(A_i) = -\sum_{i=1}^{N} \frac{|A_i|}{|X|} \log \frac{|A_i|}{|X|} \tag{2}$$

In particular, we can consider the *trivial* partition $\pi_t = \{X\}$ (consisting of the unique set X) and the *discrete* partition $\pi_d = \{\{x_1\}, \{x_2\}, \ldots, \{x_{|X|}\}\}$ (the collection of all singletons from the universe $X = \{x_1, x_2, \ldots, x_{|X|}\}$ of cardinality $|X|$). In these two particular partitions the associated entropies are $H(\pi_t) =$

0 and $H(\pi_d) = \log |X|$ and for any other partition π of the same universe X one has the following inequalities: $0 = H(\pi_t) \leq H(\pi) \leq H(\pi_d) = \log |X|$, with $H(\pi) = 0$ iff $\pi = \pi_t$ and $H(\pi) = \log |X|$ iff $\pi = \pi_d$.

Note that the entropy (2) associated with a probability distribution π assumes also the following form:

$$H(\pi) = \log |X| - \frac{1}{|X|} \sum_{i=1}^{N} |A_i| \log |A_i| \qquad (3)$$

Hence, if one introduces the *co–entropy* of the partition π defined as

$$E(\pi) := \frac{1}{|X|} \sum_{i=1}^{N} |A_i| \log |A_i| \qquad (4)$$

then the (3) leads to the identity:

$$\forall \pi, \quad H(\pi) + E(\pi) = \log |X| \qquad (5)$$

i.e., the quantity $E(\pi)$ is the "entropy" which *complements* $H(\pi)$ with respect to the constant value $\log |X|$. Let us stress, with particular regard to the fuzzy generalization discussed in the sequel, that since any $|A_i| \geq 1$ also the co–entropy is a non negative quantity whatever be the involved partition: formally, for every partition π, $E(\pi) \geq 0$. Moreover, also in this case we have the inequalities: $\forall \pi, \, 0 = E(\pi_d) \leq E(\pi) \leq E(\pi_t) = \log |X|$, with $E(\pi) = 0$ iff $\pi = \pi_d$ and $E(\pi) = \log |X|$ iff $\pi = \pi_t$.

It is possible to consider two (non–negative) discrete random variables generated by a partition π:

(**G-RV**). The *granularity random variable* $\boldsymbol{G}(\pi) = (\log |A_1|, \log |A_2|, \ldots, \log |A_N|)$, each component of which $G(A_i) := \log |A_i|$ expresses the *measure of the granularity* supported by the *granule* A_i of the partition π.

(**U-RV**). The *uncertainty random variable* $\boldsymbol{I}(\pi) = (-\log p(A_1), -\log p(A_2), \ldots, -p(A_N))$, each component of which $I(A_i) := -\log p(A_i)$ expresses (according to the general discussion of section 1) the *uncertainty measure* related to the probability of occurrence of the *event* A_i of the partition π.

The relationship between the uncertainty measure and the granularity measure of A_i is similar to the (5): $\forall A_i, \, G(A_i) + I(A_i) = \log |X|$, with $G(A_i)$ (resp., $I(A_i)$) increasing (resp., decreasing) mapping. With respect to the now introduced random variables, we have that $E(\pi) = \sum_{i=1}^{N} G(A_i) \cdot p(A_i)$, i.e., the co–entropy furnishes the *average granularity measure*, and $H(\pi) = \sum_{i=1}^{N} I(A_i) \cdot p(A_i)$, i.e., the entropy furnishes the *average uncertainty measure* related to π.

2.1 Partitions induced from information systems

Let us recall that in the context of partitions of a concrete universe, as the support structure of the Pawlak approach to rough set theory, the above co–entropy has been introduced in [4]. In particular, these considerations can be

applied to the case of a (complete) *Information System* (IS), formalized by a triple $IS := \langle X, Att, F \rangle$ consisting of a nonempty finite set X of objects, a nonempty finite set of attribute Att, and a mapping $F : X \times Att \to V$ which assigns to any object $x \in A$ the value $F(x, a)$ assumed by the attribute $a \in Att$ [5], [7], [3]. Indeed, in the IS case the partition of the universe of objects X generated by a set of attributes $\mathcal{A}$, denoted by $\pi_{\mathcal{A}}(IS)$, consists of the equivalence classes of objects which are *indistinguishable* with respect to the information furnished by the attributes in $\mathcal{A}$, formalized by the equivalence relation: $(x, y) \in R_{\mathcal{A}}$ iff $\forall a \in \mathcal{A}$, $F(x, a) = F(y, a)$.

Example 1. Consider the (complete) information system based on the finite universe $X = \{1, 2, 3, 4\}$ and finite set of attributes $Att = \{a_0, a_1, a_2\}$. The information system is given by the following *information table*:

$x \in X$	$f_{a_0}(x)$	$f_{a_1}(x)$	$f_{a_2}(x)$
1	A	G	S
2	A	R	S
3	T	G	M
4	T	G	L

The meaning of the attributes of this example could be the following: a_0 is the *shape* of the object (A as *arched* and T as *thin*), a_1 is the *color* (G as *green* and R as *red*), a_2 is the *dimension* (S as *small*, M as *medium* and L as *large*). We have the following partitions, one for each possible collection of attributes:

$$\pi(a_0) = \{\{1, 2\}, \{3, 4\}\}, \quad \pi(a_1) = \{\{1, 3, 4\}, \{2\}\},$$
$$\pi(a_2) = \{\{1, 2\}, \{3\}, \{4\}\} = \pi(a_0, a_2),$$
$$\pi(a_0, a_1) = \{\{1\}, \{2\}, \{3, 4\}\}, \quad \pi(a_1, a_2) = \{\{1\}, \{2\}, \{3\}, \{4\}\} = \pi(a_0, a_1, a_2)$$

2.2 Partition entropy on a finite measure space

Let us see now some generalizations of the partition entropy notion, making the following

Notational Convention: *From now on, if no confusion is likely, the counting measure m_c (resp., probability p_c) will be simply denoted by m (resp., p).*

This notational convention is very useful from the general point of view. Indeed, the treatment of the entropy partition performed in the previous section is formally based on the structure of the measure space $\langle X, \mathcal{E}_\pi(X), m_c \rangle$ in which the universe is constrained to the strong condition of being a *finite* set and the involved measure is the very particular counting measure assigning to any (finite) measurable subset $E \in \mathcal{E}_\pi(X)$ the corresponding cardinality $m_c(E) = |E|$.

This approach can be extended to the more general case of a measurable space with finite measure $\langle X, \mathcal{E}(X), m \rangle$ where X is a (non necessarily finite) universe, $\mathcal{E}(X)$ a fixed σ–algebra of its measurable subsets, and $m : \mathcal{E}(X) \mapsto \mathbb{R}_+$ a *finite* measure on $\mathcal{E}(X)$, i.e., such that $m(X) < \infty$. Note that this condition imply (for the standard monotonicity property of a generic measure) that for any

$E \in \mathcal{E}(X)$, $m(E) \leq m(X) < \infty$. Indeed, all the results previously proved, and all the results which we shall prove in the sequel, are formulated in such a way that they hold in this general framework, instead of the very narrow situation of a finite universe with the counting measure. It is only necessary to consider the so–called *measurable partitions* of X, i.e., *finite* (or in general, with a slight modification about convergence, *countable*) families $\pi = \{E_1, E_2, \ldots, E_N\}$ of measurable subsets of X (for any i, $E_i \in \mathcal{E}(X)$) such that: (pp-1) $m(E_i) > 0$ (strictly positiveness); (pp-2) $m(E_i \cap E_j) = 0$ (measurable disjointness); (pp-3) $m(\cup_i E_i) = \sum_{i=1}^{N} m(E_i) = m(X)$ (normalization). Of course, any standard partition π of X is also a measure partition, i.e., we have an enrichment of the usual notion of partition.

Given a probability partition π, the corresponding vector $\boldsymbol{p}(\pi) = \left(\frac{m(E_1)}{m(X)}, \frac{m(E_2)}{m(X)}, \ldots, \frac{m(E_N)}{m(X)} \right)$ is a probability distribution. More generally, we have the probability space $\langle X, \mathcal{E}(X), p_m \rangle$, where $p_m(E) = \frac{m(E)}{m(X)}$ is a probability measure on the σ–algebra $\mathcal{E}(X)$ generated by the finite measure m.

As examples of this general situation, let us mention any Lebesgue measurable subset X of $\mathbb{R}^n$ of finite Lebesgue measure $\mu(X) < \infty$. For instance, a bounded interval of $\mathbb{R}$, a bounded rectangle and a circle of $\mathbb{R}^2$, a bounded parallelepiped or a sphere in $\mathbb{R}^3$, and so on.

Let us stress that also in the case of a *finite* universe the just defined notion of probability partition leads to an enrichment of the usual family of partitions.

Example 2. Let us consider a *biased die* modelled by a probability space $\langle X, \mathcal{A}(X), p \rangle$ with $X = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A}(X) = \mathcal{P}(X)$, and the probability function generated on $\mathcal{A}(X)$ by the probabilities defined for any elementary event $\{i\}$ of the involved universe X by the probability distribution $\boldsymbol{p} = (p(\{1\}) = 2/6, p(\{2\}) = 1/6, p(\{3\}) = 1/6, p(\{4\}) = 1/6, p(\{5\}) = 1/6, p(\{6\}) = 0)$. The probability of the generic subset (event) A of X is given by the rule: $p(A) = \sum_{i \in A} p(\{i\})$. We have in particular that $p(X) = \sum_{i \in X} p(\{i\}) = 1$. Then the families of events $\pi_1 = \{A_1 = \{1, 2, 6\}, A_2 = \{4, 6\}, A_3 = \{5, 6\}\}$ and $\pi_2 = \{B_1 = \{1, 2, 3\}, B_2 = \{4, 5\}\}$ are probability partitions which are not standard. The former π_1 is a covering of X and the latter π_2 is not a covering, but its two subsets are disjoint.

2.3 Partitions as identity resolutions by crisp sets (sharp granulations)

In this section we show that any partition can be identify with an "*identity resolution* by crisp sets. To be precise, given a partition $\pi = \{A_1, A_2, \ldots, A_N\}$), if one introduces the characteristic functional $\chi_{A_i} : X \mapsto \{0, 1\}$ of any set A_i defined for any point $x \in X$ as $\chi_{A_i}(x) = 1$ if $x \in A_i$ and $= 0$ otherwise, then the collection of characteristic functionals $\mathcal{C}(\pi) := \{\chi_{A_1}, \chi_{A_2}, \ldots, \chi_{A_N}\}$ associated with the partition π is a *crisp (sharp) identity resolution*, i.e., a family of crisp sets such that the following property holds: $\forall x \in X$, $\sum_{i=1}^{N} \chi_{A_i}(x) = 1$. Denoting

by $\mathbf{1}$ the identity mapping assigning to any point $x \in X$ the value $\mathbf{1}(x) = 1$, we can also say that the family of crisp sets $\mathcal{C}(\pi)$ satisfies the *functional crisp identity resolution* condition:

$$\sum_{i=1}^{N} \chi_{A_i} = \mathbf{1} \tag{6}$$

Of course, partitions and identity resolutions can be identified by the one-to-one and onto correspondence

$$\pi = \{A_1, A_2, \ldots, A_N\} \longleftrightarrow \mathcal{C}(\pi) := \{\chi_{A_1}, \chi_{A_2}, \ldots, \chi_{A_N}\} \tag{7}$$

The condition (6) defines the identity resolution by crisp sets $\mathcal{C}(\pi) = \{\chi_{A_i} \in \{0,1\}^X : i = 1, 2, \ldots, N\}$ as a *crisp granulation* of the universe X, in which any crisp set χ_{A_i} is a *granule*. Then, the *counting measure* of the *elementary event* A_i by its *crisp granule* representation χ_{A_i} is given by

$$m(A_i) = \sum_{x \in X} \chi_{A_i}(x) = |A_i| \tag{8}$$

The probabilities associated to any event A_i can also be expressed as a probability of the corresponding crisp granule χ_{A_i} by the equation

$$p(A_i) = \frac{1}{m(X)} m(A_i) = \frac{1}{m(X)} \sum_{x \in X} \chi_{A_i}(x) \tag{9}$$

and the entropy (2) and co–entropy (4) generated by π are given now by

$$H(\pi) = \log m(X) - \frac{1}{m(X)} \sum_{i=1}^{N} m(A_i) \log m(A_i) \tag{10a}$$

$$E(\pi) = \frac{1}{m(X)} \sum_{i=1}^{N} m(A_i) \log m(A_i) \tag{10b}$$

with the standard result $\forall \pi, \quad H(\pi) + E(\pi) = \log m(X)$.

3 Fuzzy (Unsharp) Granulations

An immediate generalization of crisp identity resolution (as a finite collection of crisp sets $\mathcal{C} := \{\chi_i \in \{0,1\}^X : i = 1, 2, \ldots, N\}$ whose sum is, according to (6), the identity mapping $\sum_{i=1}^{N} \chi_i = \mathbf{1}$) is the notion of *fuzzy identity resolution* as a finite collection of fuzzy sets $\mathcal{F} := \{\omega_i \in [0,1]^X : i = 1, 2, \ldots, N\}$ such that the functional identity resolution condition $\sum_{i=1}^{N} \omega_i = \mathbf{1}$ holds. Generalizing the (8), the measure of a generic fuzzy set $\omega \in [0,1]^X$ of a *finite* universe X can be defined as follows

$$m(\omega) := \sum_{x \in X} \omega(x) \tag{11}$$

Definition 1. *A* fuzzy *(also* unsharp*) granulation of the universe X is defined as a collection of fuzzy sets $\mathcal{F} = \{\omega_i \in [0,1]^X : i = 1, 2, \ldots, N\}$, whose elements ω_i are said to be* fuzzy granules*, under the condition of* total measure *$M(\mathcal{F}) := \sum_{i=1}^{N} m(\omega_i) = |X|$.*

Thus, the condition of fuzzy granulation is dependent from the measure $m(\omega_i)$ of each *fuzzy granule* $\omega_i \in \mathcal{F}$ given by (11), provided its total measure is the cardinality of the universe. Let us stress that in practical applications some further regularity conditions are usually (hiddenly, in the sense of non explicitly formalized) involved. Let us quote two of them which in this paper are tacitly assumed.

Non–redundancy of the granulation, formally expressed by the fact that if in a fuzzy granulation $\mathcal{F}$ two fuzzy granules $\omega_i, \omega_j \in \mathcal{F}$ are such that $\omega_i \leq \omega_j$ (in the pointwise ordering, i.e., for every $x \in X$ one has $\omega_i(x) \leq \omega_j(x)$), then $\omega_i = \omega_j$.

Covering of the universe X, i.e., for any point $x \in X$ there must exists at least a fuzzy granule $\omega_i \in \mathcal{F}$ such that $\omega_i(x) \neq 0$.

Let us note that any fuzzy identity resolution, $\mathcal{F} = \{\omega_i \in \{0,1\}^X : i = 1, 2, \ldots, N\}$ is necessarily a fuzzy granulation since $\sum_{i=1}^{N} m(\omega_i) = \sum_{i=1}^{N} \sum_{x \in X} \omega_i(x) = \sum_{x \in X} \sum_{i=1}^{N} \omega_i(x) = |X|$. The vice versa in general is not true as the following example shows.

Example 3. In the finite universe $X = \{1, 2, 3, 4, 5\}$, let us consider the fuzzy granulation consisting of the two fuzzy sets ω_i, $i = 1, 2$, defined by the table

	1	2	3	4	5	$m(\omega_i)$
ω_1	1	1/2	1/2	1/2	0	5/2
ω_2	1/2	0	1/2	1	1/2	5/2

Then $m(\omega_1) = m(\omega_2) = 5/2$, from which $\sum_{i=1,2} m(\omega_i) = 5 = |X|$. But, for instance, $\omega_1(4) + \omega_2(4) = 3/2 \neq 1$.

Moreover, any fuzzy identity resolution is a covering fuzzy granulation, but in general there is no certainty about the non–redundancy condition.

Example 4. In the case of the universe $X = \{1, 2, 3\}$ the following fuzzy identity resolution of two fuzzy sets furnishes a fuzzy granulation of X which is redundant.

	1	2	3	$m(\omega_i)$
ω_1	1/2	1	1/2	2
ω_2	1/2	0	1/2	1

Indeed, $m(\omega_1) + m(\omega_2) = 3$, but $\omega_2(x) \leq \omega_1(x)$ whatever be x.

3.1 Entropy and (possible negative) co–entropy for fuzzy granulation

In the case of a fuzzy granulation, for any of its fuzzy granule ω_i it is possible to assign the non–negative number (and compare with (9)):

$$p(\omega_i) = \frac{1}{M(\mathcal{F})} m(\omega_i) = \frac{1}{M(\mathcal{F})} \sum_{x \in X} \omega_i(x) \tag{12}$$

whose collection $\boldsymbol{p}(\mathcal{F}) = (p(\omega_1), p(\omega_2), \ldots, p(\omega_N))$ is a *probability distribution* since trivially: (pd-1) for any i it is $p(\omega_i) \geq 0$ and (pd-2) $\sum_{i=1}^{N} p(\omega_i) = 1$. The *entropy* of the fuzzy granulation $\mathcal{F}$ is then the one generated by this probability distribution, which as usual is given by the real non–negative quantity

$$0 \leq H(\mathcal{F}) = -\sum_{i=1}^{N} p(\omega_i) \log p(\omega_i) \leq \log N \tag{13}$$

Trivially, by (12) and recalling that the total measure of the fuzzy covering $\mathcal{F}$ is $M(\mathcal{F}) = |X|$, one gets that

$$H(\mathcal{F}) = -\frac{1}{M(\mathcal{F})} \sum_{i=1}^{N} m(\omega_i) \log \frac{m(\omega_i)}{M(\mathcal{F})} \tag{14}$$

and so also in this case we can introduce the *co–entropy* of the fuzzy granulation $\mathcal{F}$ as the quantity

$$E(\mathcal{F}) = \frac{1}{M(\mathcal{F})} \sum_{i=1}^{N} m(\omega_i) \log m(\omega_i) \tag{15}$$

obtaining from (14) the following identity which is true whatever be the fuzzy granulation $\mathcal{F}$:

$$H(\mathcal{F}) + E(\mathcal{F}) = \log M(\mathcal{F}) = \log |X| \tag{16}$$

This identity is an extension to fuzzy granulations of the identity (5) previously seen in the case of partitions (crisp granulations). Also in this case the "co–entropy" $E(\mathcal{F})$ complements the original entropy $H(\mathcal{F})$ with respect to the constant quantity $\log |X|$, invariant relatively to the choice of the granulation $\mathcal{F}$. This co–entropy refers to the *measure distribution* $\boldsymbol{m}(\mathcal{F}) = (m(\omega_1), m(\omega_2), \ldots, m(\omega_N))$ for which the following hold: (md-f1) every $m(\omega_i) \geq 0$; (md-f2) its *total measure* is $M(\mathcal{F}) := \sum_{i=1}^{N} m(\omega_i) = |X|$. Of course, the entropy of a fuzzy granulation, from (13), is always non–negative, but in the present fuzzy case notwithstanding the expected link expressed by (16), the co–entropy could be negative.

Example 5. In the universe $X = \{1, 2, 3\}$ let is consider the fuzzy granulation consisting of the fuzzy sets defined according to the following tabular represen-

tation:

	1	2	3	$m(\omega_i)$
ω_1	1/2	0	0	1/2
ω_2	1/2	1/3	0	5/6
ω_3	0	1/3	1	4/3
ω_4	0	1/3	0	1/3

Of course, this is a fuzzy identity resolution (and so also a fuzzy granulation) since trivially $\sum_{i=1}^{4} \omega_i(x) = 1$ for every point $x = 1, 2, 3$. But in this example the entropy is $H \cong 1.8163$, whereas the co–entropy is negative $E \cong -0.2314$ with $H + E \cong 1.5850 \cong \log 3$.

As shown by this example, the possible negativity of co–entropy (15) rises from the fact that some of the measures $m(\omega_i)$ could be number in the real unit interval $[0,1]$ producing in this way a $\log m(\omega_i)$ term which is negative. In the case of a fuzzy granulation $\mathcal{F}$ it is possible to consider the two following discrete random variables:

(**FG-RV**). The *fuzzy granularity random variable* $\boldsymbol{G}(\mathcal{F}) = (\log m(\omega_1), \log m(\omega_2), \ldots, \log m(\omega_N))$, each component of which $G(\omega_i) := \log m(\omega_i)$ expresses the *granularity measure* supported by the *fuzzy granule* ω_i of the fuzzy granulation $\mathcal{F}$. Note that some of these fuzzy granules could have negative measure, precisely under the condition $G(\omega_i) < 1$.

(**FU-RV**). The non–negative *fuzzy uncertainty random variable* $\boldsymbol{I}(\mathcal{F}) = (-\log p(\omega_1), -\log p(\omega_2), \ldots, -p(\omega_N))$, each component of which $I(\omega_i) := -\log p(\omega_i) = \log M(\mathcal{F}) - \log m(\omega_i)$ expresses (according to the general discussion of section 1) the *uncertainty measure* related to the probability of occurrence of the *fuzzy event* ω_i of the fuzzy granulation $\mathcal{F}$.

The relationship between these uncertainty and granularity measures of ω_i is now (compare with the (16)): $\forall \omega_i$, $G(\omega_i) + I(\omega_i) = \log |X|$. With respect to the now introduced random variables, we have that $E(\mathcal{F}) = \sum_{i=1}^{N} G(\omega_i) \cdot p(\omega_i)$, i.e., it is the *average fuzzy granularity measure*, and $H(\mathcal{F}) = sum_{i=1}^{N} I(\omega_i) \cdot p(\omega_i)$, i.e., it is the *average fuzzy uncertainty measure* related to the fuzzy granulation.

3.2 A normalized non–negative co–entropy for fuzzy granulation

In order to avoid the previously stressed negativity of co–entropy of a fuzzy granulation $\mathcal{F}$, due to the fact that the measure distribution $\boldsymbol{m}(\mathcal{F}) = (m(\omega_1), m(\omega_2), \ldots, m(\omega_N))$ some of its components could be less that 1, it is possible to introduce the minimum measure $\hbar(\mathcal{F}) := \min\{m(\omega_1), m(\omega_2), \ldots, m(\omega_N)\}$, and then to construct the new fuzzy granulation $\mathcal{F}_\hbar = \mathcal{F}/\hbar := (\omega_1/\hbar(\mathcal{F}), \omega_2/\hbar(\mathcal{F}), \ldots, \omega_N/\hbar(\mathcal{F}))$, where the generic component is $\omega_i/\hbar(\mathcal{F}) \geq 1$. The measure distribution corresponding to the new fuzzy granulation $\mathcal{F}_\hbar$ is $\boldsymbol{m}_\hbar(\mathcal{F}) = (m(\omega_1)/\hbar(\mathcal{F}), m(\omega_2)/\hbar(\mathcal{F}), \ldots, m(\omega_N)/\hbar(\mathcal{F}))$, whose total measure is $M_\hbar(\mathcal{F}) = \sum_{i=1}^{N} m(\omega_i)/\hbar(\mathcal{F}) = |X|/\hbar(\mathcal{F})$, and with generic component denoted by $m_\hbar(\omega_i) := m(\omega_i)/\hbar(\mathcal{F})$. The probability distribution generated by the normalization of the new measure distribution $\boldsymbol{m}_\hbar(\mathcal{F})$ by its total measure $M_\hbar(\mathcal{F})$ is $\boldsymbol{p}_\hbar :=$

$m_{\hbar}(\mathcal{F})/M_{\hbar}(\mathcal{F}) = m(\mathcal{F})/|X| = p$, i.e., the probability distribution does not change as to this change of the measure distribution. As a consequence of this result, also the entropy does not change: $H(\mathcal{F}_{\hbar}) = H(\mathcal{F})$. It is the co–entropy which strongly changes (compare with (15)):

$$E(\mathcal{F}_{\hbar}) = \frac{1}{M_{\hbar}(\mathcal{F})} \sum_{i=1}^{N} m_{\hbar}(\omega_i) \cdot \log m_{\hbar}(\omega_i) \tag{17}$$

In particular $E(\mathcal{F}_{\hbar}) = E(\mathcal{F}) - \log \hbar(\mathcal{F})$, and so with respect to the new quantities we have that (and compare with (16)):

$$H_{\hbar}(\mathcal{F}) + E_{\hbar}(\mathcal{F}) = \log \frac{M(\mathcal{F})}{\hbar(\mathcal{F})} = \log \frac{|X|}{\hbar(\mathcal{F})} \tag{18}$$

In particular, from $[H_{\hbar}(\mathcal{F}) + \log \hbar(\mathcal{F})] + E_{\hbar}(\mathcal{F}) = \log M(\mathcal{F})$, we can introduce a new entropy for fuzzy granulation, $H'_{\hbar}(\mathcal{F}) = H_{\hbar}(\mathcal{F}) + \log \hbar(\mathcal{F})$, for which trivially one has the expected "invariance" $H'_{\hbar}(\mathcal{F}) + E_{\hbar}(\mathcal{F}) = \log M(\mathcal{F}) = \log |X|$. Explicitly it turns out that this new entropy has the form (and compare with (14)):

$$H'_{\hbar}(\mathcal{F}) = -\frac{1}{M_{\hbar}(\mathcal{F})} \sum_{i=1}^{N} m_{\hbar}(\omega_i) \cdot \log \frac{m_{\hbar}(\omega_i)}{M_{\hbar}(\mathcal{F})} \frac{1}{\hbar(\mathcal{F})} \tag{19}$$

4 Conclusions and open problems

The extension of the crisp granulation notion to the fuzzy case is investigated. In particular the two usual *measures* of average granulation (co–entropy) and average uncertainty (entropy) is deeply treated, eliminating a first drawback of possible negativity of the co–entropy.

As an open problem it remains to study the behavior of the fuzzy co–entropy and entropy with respect to the monotonicity.

References

1. R. B. Ash, *Information theory*, Dover Publications, New York, 1990, (originally published by John Wiley & Sons, New York, 1965).
2. R. V. L. Hartley, *Transmission of information*, The Bell System Technical Journal **7** (1928), 535–563.
3. J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, *Rough sets: A tutorial*, Rough Fuzzy Hybridization (S. Pal and A. Skowron, eds.), Springer–Verlag, Singapore, 1999, pp. 3–98.
4. J. Liang and Z. Shi, *The information entropy, rough entropy and knowledge granulation in rough set theory*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **12** (2004), 37–46.
5. Z. Pawlak, *Information systems - theoretical foundations*, Information Systems **6** (1981), 205–218.
6. _____, *Rough sets*, Int. J. Inform. Comput. Sci. **11** (1982), 341–356.
7. _____, *Rough sets: Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Dordrecht, 1991.