ENTROPIES AND CO–ENTROPIES OF COVERINGS WITH APPLICATION TO INCOMPLETE INFORMATION SYSTEMS

D. BIANUCCI, G. CATTANEO, AND D. CIUCCI

Abstract. Different generalizations to the case of coverings of the standard approach to entropy applied to partitions of a finite universe X are explored. In the first approach any covering is represented by an identity resolution of fuzzy sets on X and a corresponding probability distribution with associated entropy is defined. A second approach is based on a probability distribution generated by the covering normalizing the standard counting measure. Finally, the extension to a generic covering of the Liang–Xu approach to entropy is investigated, both from the "global" and the "local" point of view. For each of these three possible entropies the complementary entropy (or co–entropy) is defined showing in particular that the Liang–Xu entropy is a co–entropy.

1. Introduction: the link between Information and Rough Theories

The notion of partition of a (finite) set, the *universe* of the discourse, plays a fundamental role both in Pawlak rough set theory [Paw82] and in Shannon information theory [Sha48]. Recently, a certain interest in using the entropy notion typical of information systems in the framework of rough set theory can be found in literature, either in the case of the universe partition generated by a complete information system (see for instance [Wie99, LS04], and [Sle02] for applications to the reduction of attributes) or in the case of the universe covering generated by an incomplete information system (see for instance [LX00, HHZ04]). Let us make a brief introduction about these subjects and their possible relationship.

The original Pawlak approach to rough sets can be summarized in a pair $\langle X, \pi \rangle$ consisting of a (finite) set X, the *universe* of objects, and a partition $\pi = \{A_1, A_2, \ldots, A_N\}$ of X whose elements A_j are called *elementary sets* or also *granules of knowledge*; in the sequel we denote by $\pi(x)$ the granule which contains the point $x \in X$. The partition π generates the binary equivalence relation $\mathcal{R}(\pi)$ on X defined by the law: $(x, y) \in \mathcal{R}(\pi)$ iff $\exists A_i \in \pi$: $x \in A_i$ and $y \in A_i$. This equivalence relation is interpreted as an *indiscernibility* relation about the objects of the universe in the sense that if two objects x and y of X belong to the same subset A_i of the partition, then they cannot be discerned with respect to π. Elementary sets are then considered as supports of knowledge and so, if the knowledge about objects is concentrated in π, any other subset H of objects from the universe can be approximated from the bottom and from the top by the two *lower* and *upper approximations* defined respectively as:

$$l(H) := \cup\{A_i \in \pi : A_i \subseteq H\} \quad \text{and} \quad u(H) := \cup\{A_j \in \pi : H \cap A_j \neq \emptyset\} \tag{1}$$

Trivially, the following order chain holds: $l(H) \subseteq H \subseteq u(H)$ and since both $l(H)$ and $u(H)$ are set theoretic union of elementary sets from π, in rough set theory it is customary to define as *definable set* any possible set theoretic union of elementary sets. The collection $\mathcal{E}_\pi(X)$ of all definable subsets of X plus the empty set constitute a topology of *clopen* sets, or from another point of view a σ–algebra of *measurable* sets; in this "measure" context subsets from $\mathcal{E}_\pi(X)$ are also called *events* and elements from π *elementary events.* All the other elements of the power set $\mathcal{P}(X)$ of X are subsets of objects which can be *approximated* by definable sets according to (1), where the *rough approximation* of H is the pair of events $r_\pi(H) := \langle l(H), u(H)\rangle$. Let us recall that a subset E of X is said to be *sharp* (or *crisp*) iff $l(E) = u(E)$, and this happens iff $E \in \mathcal{E}_\pi(X)$, i.e., it is an event.

From the information point of view, to any elementary set A_j of a partition π it is possible to associate the corresponding probability of occurrence $p(A_j) = \frac{|A_j|}{|X|} \geq 0$, characterized by the property that the $p(A_j)$ constitute a *probability distribution* since $\sum_{j=1}^{N} p(A_j) = 1$. Let us recall (and see section 2) that in information theory the non negative real number $I(A_j) = -\log p(A_j)$ "measures" the uncertainty due the knowledge of a probabilistic information (in this paper all the involved logarithms are to the base 2). In this way, any partition π generates two vectors: the *uncertainty vector* $\vec{I}(\pi) := (I(A_1), I(A_2), \ldots, I(A_N))$ and the *probability distribution vector* $\vec{p}(\pi) := (p(A_1), p(A_2), \ldots, p(A_N))$. The family of pairs $\boldsymbol{\pi} := \{(I(A_i), p(A_i) : i = 1, 2, \ldots, N\}$ is the *probability scheme* generated by the partition π, represented by the *finite scheme matrix* [Khi57]:

$$\boldsymbol{\pi} := \begin{bmatrix} I(A_1) & I(A_2) & \ldots & I(A_N) \\ p(A_1) & p(A_2) & \ldots & p(A_n) \end{bmatrix} \tag{2}$$

Thus, the *average uncertainty* of this probability scheme generated by the partition π is given by the *entropy* $H(\pi) := \sum_{j=1}^{N} I(A_j)\, p(A_j) = -\sum_{j=1}^{N} p(A_j) \log p(A_j)$.

1.1. Application of information theory to rough set theory.

A possible application of rough set and information theory, both based on the mathematical notion of partition, could be the field of *information systems* (as introduced by Pawlak in [Paw81]), also called *knowledge representation systems* (according to Vakarelov [Vak91]), and formalized by an *information table* (see also [KPPS99, Paw91]). The rows of the table are labelled by *objects* $x_1, x_2, \ldots, x_N \in X$ and the columns by *attributes* $a_1, a_2, \ldots, a_M \in Att$. The table entry corresponding to the object $x \in X$ and the attribute $a \in Att$ is formalized by $F(x, a)$, which is a particular *attribute-value*. In this way one can denote an information system as a triple $\langle X, Att, F\rangle$ where F is a mapping defined on the set $X \times Att$ of all object–attribute pairs (x, a), which assigns to each of these pairs a *value* $F(x, a)$. The map F provides a kind of *observation* of the properties that can be taken into consideration for each object in the universe. More precisely, given an attribute $a \in Att$, one can define the set $val(a) := \{\alpha : \exists x \in X \ s.t.\ F(x, a) = \alpha\}$ containing all the specific possible values of a. On this basis, the observation of the attribute $a \in Att$ on an object $x \in X$ yields the value $F(x, a) \in val(a)$. Thus, we can introduce a family of (surjective) mappings $f_a : X \to val(a)$, each of which is bijectively depending from the observed attribute a and defined by the correspondence $x \mapsto f_a(x) := F(x, a)$. Let $val = \cup_{a\in Att} val(a)$ be the global set of possible values of the information system, then each attribute a can be identified with the mapping $f_a \in val^X$ and so, introducing the collection of all such mappings $Att(X) := \{f_a \in val^X : a \in Att\}$ in which Att plays the role of index set, an information system can be formalized also as a structure $\langle X, Att(X)\rangle$. Quoting from [SW01]: "*Data analysis* concerns features labelling known cases with specific values. A sample of data can take the form of an *information system* $\langle X, Att\rangle$, where each attribute $a \in Att$ is identified with a function $f_a : X \to val(a)$ from the universe of objects X into the set $val(a)$ of all possible values on a."

Any fixed attribute $a \in Att$, with set of values $val(a) = \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$, generates a partition of the universe of objects $\pi(a) = \{f^{-1}(\alpha_1), f^{-1}(\alpha_2), \ldots, f^{-1}(\alpha_N)\}$, where the generic elementary event of $\pi(a)$ is $f^{-1}(\alpha_i) := \{x \in X : f_a(x) = \alpha_i\}$, i.e., collection of all objects with respect to which the attribute a assumes the fixed value α_i. The pair (a, α_i) is interpreted as the *elementary proposition* "an observation of a yields the result α_i" and $A_i := f^{-1}(\alpha_i)$ is the *event* which tests the proposition (a, α_i), in the sense that it is constituted by all objects with respect to which the proposition (a, α_i) is "true" ($x \in A_i$ iff $f_a(x) = \alpha_i$). The event A_i is then the equivalence class (also denoted by $[a, \alpha_i]$), of all objects on which the attribute a assumes the value α_i, and $p(A_i) = p([a, \alpha_i]) = \frac{|A_i|}{|X|}$ is the probability of occurrence of this event, or equivalently of the proposition (a, α_i). The average uncertainty of the partition generated by the attribute a is then quantified by the entropy $H(a) = -\sum_{i=1}^{N} \frac{|A_i|}{|X|} \log \frac{|A_i|}{|X|} = \log |X| - \frac{1}{|X|} \sum_{i=1}^{N} |A_i| \log |A_i|$.

If we consider a set $\mathcal{A}$ consisting of two attributes $a \in Att$ and $b \in Att$ with corresponding set of values $val(a) = \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ and $val(b) = \{\beta_1, \beta_2, \ldots, \beta_M\}$ (the generalization to a generic collection of attributes will be discussed later), then it is possible to define the mapping $f_{a,b} : X \mapsto val(a, b) := val(a) \times val(b)$ which assigns to any object x the "value" $f_{a,b}(x) := (f_a(x), f_b(x))$. In this case we can consider the pair $(a, b) \in Att^2$ as a single attribute of the new information system $\langle X, Att^2, \{f_{a,b} : a, b \in Att\}\rangle$, always based on the original universe X. The partition generated by the attribute (a, b) is then the collection of subsets of X, $\pi(a, b) =$

$\{f_{a,b}^{-1}(\alpha_i,\beta_j) : \alpha_i \in val(a) \text{ and } \beta_j \in val(b)\}$, where the generic elementary event of the partition $\pi(a,b)$ is the subset of the universe

$$f_{a,b}^{-1}(\alpha_i,\beta_j) := \{x \in X : f_{a,b}(x) = (\alpha_i,\beta_j)\} = f_a^{-1}(\alpha_i) \cap f_b^{-1}(\beta_j)$$

Hence, adopting the notation of $[(a,\alpha_i) \ \& \ (b,\beta_j)]$ to denote the elementary event $f^{-1}(\alpha_i,\beta_j)$ we have that $[(a,\alpha_i) \ \& \ (b,\beta_j)] = [a,\alpha_i] \cap [b,\beta_j]$. If $(a,\alpha_i) \ \& \ (b,\beta_j)$ is interpreted as the *conjunction* "a test of a yields the result α_i *and* of b yields the result β_j" (i.e., & represents the logical connective "and" between propositions), then this result says that the set of objects in which this proposition is verified is just the set of objects in which simultaneously "a yields α_i" *and* "b yields β_j".

On the other hand, making use of the notations $C_{i,j} := f^{-1}(\alpha_i,\beta_j)$, $A_i = f_a^{-1}(\alpha_i)$ and $B_j = f_b^{-1}(\beta_j)$ we can reformulate the previous result as $C_{i,j} = A_i \cap B_j$. In other words, elementary events from the partition $\pi(a,b)$ are obtained as set theoretic intersection of elementary events from the partitions $\pi(a)$ and $\pi(b)$. This fact is denoted by $\pi(a,b) = \pi(a) \cdot \pi(b)$. This result can be easily extended to the case of arbitrary families of attributes. Indeed, let $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ be such a family of attributes from an information system. Then, it is possible to define the partition $\pi(\mathcal{A}) = \pi(a_1, a_2, \ldots, a_k) = \pi(a_1) \cdot \pi(a_2) \cdot \ldots \cdot \pi(a_k) := \{A_i \cap B_j \cap \ldots \cap K_p : A_i \in \pi(a_1), B_j \in \pi(a_2), \ldots, K_p \in \pi(a_k)\}$. If now one considers another family of attributes $\mathcal{B} = \{b_1, b_2, \ldots, b_h\}$ then

$$\pi(\mathcal{A} \cup \mathcal{B}) := \pi(a_1, \ldots, a_k, b_1, \ldots, b_h) = \pi(\mathcal{A}) \cdot \pi(\mathcal{B}) \tag{3}$$

Let us note that the equivalence relation $\mathcal{R}_\mathcal{A}$ induced on the universe X from the partition $\pi(\mathcal{A})$ is defined as follows:

$$(x,y) \in \mathcal{R}_\mathcal{A} \quad \text{iff} \quad \forall a \in \mathcal{A} : f_a(x) = f_a(y) \tag{4}$$

The generic equivalence class generated by an object $x \in X$ is then $\pi_\mathcal{A}(x) = \{y \in X : \forall a \in \mathcal{A},\ f_a(y) = f_a(x)\}$, i.e., collection of all objects y which cannot be *distinguished* from x relatively to the information furnished by the attributes collected in $\mathcal{A}$.

1.2. **Extension to incomplete information systems.** In this paper we discuss the extension of the just described approach based on partitions from (complete) information systems, to the case of information systems which are not complete. An *incomplete* information system is formalized as a triple $\langle X, Att, F\rangle$ in which as usual X is the finite set of objects, Att the finite set of attributes, and F a mapping *partially* defined on a subset $\mathcal{D}(F)$ of $X \times Att$, but under the *non–redundancy* condition about objects: fore every object $x \in X$ at least an attribute $a \in Att$ exists such that $(x,a) \in \mathcal{D}(X)$. Of course, in an incomplete information system the mapping representation of an attribute a is partially defined on a subset X_a of X. Precisely, any attribute a can be represented as a partially defined mapping $f_a : X_a \mapsto val(a)$, with definition domain $X_a := \{x \in X : (x,a) \in \mathcal{D}(F)\}$, which in general is a subset of X. The non–redundancy condition assures that $\bigcup_{a\in Att} X_a = X$, i.e., the covering condition of X.

Also in the case of incomplete information systems, the pair $(a,\alpha_i) \in Att \times val(a)$ represents the elementary proposition "the test of a yields the value α_i" and the subset of the universe $A_i = f_a^{-1}(\alpha_i) = \{x \in X_a : f(x) = \alpha_i\}$ the elementary event of all objects for which the proposition (a,α_i) is "true". Trivially, $\bigcup_{\alpha_i\in val(a)} f_a^{-1}(\alpha_i) = X_a$, and so $\bigcup_{\substack{a\in Att\\ \alpha_i\in val(a)}} f_a^{-1}(\alpha_i) = X$, i.e., the collection of *all* elementary events $\{f_a^{-1}(\alpha_i) : a \in Att \text{ and } \alpha_i \in val(a)\}$ generated by all possible attributes from an incomplete information system is a *covering* of the universe X which is *not* a partition.

1.2.1. *The covering induced on an incomplete information system by a similarity relation.* On the other hand, for any attribute a let us denote by $f_a(x) = *$ the fact that the value corresponding to an object $x \in X \setminus X_a$ is unknown. Let us denote by $val^*(a)$ the set of all possible values of the attribute a plus the symbol $*$, then on the collection of all objects X we define for a fixed set of attributes $\mathcal{A} \subseteq Att$ the *similarity* (reflexive and symmetric, but in general non–transitive) binary relation:

$$\text{(5)}\quad \text{Let } x,y\in X, \text{ then } \quad (x,y)\in\mathcal{S}_{\mathcal{A}} \quad \text{iff} \quad \forall a_i\in\mathcal{A}, \\ \text{either} \quad f_{a_i}(x)=f_{a_i}(y) \quad \text{or} \quad f_{a_i}(x)=* \quad \text{or} \quad f_{a_i}(y)=*$$

This similarity relation is a generalization of the equivalence relation (4) introduced for complete information systems. Quoting [YLLL94] "The requirement of an equivalence relation seems to be a stringent condition that may limit the application domain of the standard rough set model." In literature one can find some generalization of a "discernibility" relation (i.e., equivalence relation) such as the "compatibility" relation introduced by Zakowski [Zak83], the "weak discernibility relation" of Vakarelov [Vak91], which is a compatibility relation (for a use of compatibility relations in logics see for instance [Orl85]).
The (nonempty) *similarity class* generated by an object x is defined as the collection of all the elements which are indistinguishable from x with respect to the similarity relation $\mathcal{S}_{\mathcal{A}}$:

$$s_{\mathcal{A}}(x) := \{y\in X : (x,y)\in\mathcal{S}_{\mathcal{A}}\}$$

The collection of all similarity classes is a covering of the universe, since $\cup_x s_{\mathcal{A}}(x)=X$, but it is not assured that these classes are mutually disjoint.

2. Information function and entropy of discrete probability distributions

In this section we briefly discuss the *abstract* approach to information theory. Abstract in the sense that it does not make any reference to a concrete universe X, but only to 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. First of all, let us introduce as *information function* the mapping $I:(0,1]\mapsto\mathbb{R}$ assigning to any *probability value* $p\in(0,1]$ the real number $I(p) := -\log(p)$ (also called the Hartley *measure of uncertainty* [Har28]). This is the unique, up to an arbitrary positive constant multiplier, function satisfying the conditions: (1) it is non-negative; (2) it satisfies the so–called *Cauchy functional condition* $I(p_1\cdot p_2)=I(p_1)+I(p_2)$; (3) it is continuous; (4) it is non–trivial ($\exists p_0\in(0,1]$ s.t. $I(p_0)\neq 0$). The information function is considered as a measure 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. Coherently, the lower is the probability, the greater is the corresponding uncertainty measure.

A length N *probability distribution* is a vector $\vec{p}=(p_1,p_2,\dots,p_N)$ in which: (i) every $p_1\geq 0$ and (ii) $\sum_{i=1}^{n} p_i = 1$. A length N *random variable* is a vector $\vec{a}=(a_1,a_2,\dots,a_N)\in\mathbb{R}^N$. For a fixed length N random variable $\vec{a}$ and a length N probability distribution $\vec{p}$ the pair $(\vec{a},\vec{p}) := \{(a_i,p_i) : i=1,2,\dots,N\}$ constitutes a *probability scheme*, abstraction of the probability scheme represented in equation (2) relative to a partition of a concrete universe X. In this abstract context of probability schemes, the numbers a_i are interpreted as *possible values* of a *discrete random variable* denoted by α and the quantities p_i as the probability of occurrence of the event $\{\alpha=a_i\}$ (thus, p_i can be considered as a simplified notation of $p(a_i)$) (see [Ash90, p.5]). Hence, the *average* (or *mean*) value of the random variable $\vec{a}$ with respect to a probability distribution $\vec{p}$ is given by $Av(\vec{a},\vec{p})=\sum_{i=1}^{N} a_i\cdot p_i$.

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

$$H(\vec{p}) = -\sum_{i=1}^{N} p_i\log p_i \tag{6}$$

Since the information $I(p)$ of a probability value p has been interpreted as a measure of the uncertainty due to the knowledge of this probability, the entropy of a probability distribution $\vec{p}$ can be considered as a quantity which in a reasonable way measures the amount of uncertainty associated with this distribution, expressed as the mean value of the corresponding information variable $\vec{I}[\vec{p}]$. Indeed, given a probability distribution $\vec{p} = (p_1, p_2, \ldots, p_N)$, its entropy $H(\vec{p}) = 0$ iff one of the numbers $p_1, p_2, \ldots, p_N$ is one and all the others are zero, and this is just the case where the result of the experiment can be predicted beforehand with complete certainty, so that there is no uncertainty as to its outcome. These probability distributions will be denoted by the conventional symbol $\vec{p}_k = (\delta_k^i)_{i=1,2,\ldots,N}$, where δ_k^i is the Kronecker delta centered in k. In all other cases the entropy is a (strongly) positive number upper bounded by $\log N$. On the other hand, given a probability distribution $\vec{p} = (p_1, \ldots, p_N)$ the entropy $H(\vec{p}) = \log N$ iff $p_i = \frac{1}{N}$ for all $i = 1, \ldots, N$ (uniform probability distribution $\vec{p}_u = (1/N, 1/N, \ldots, 1/N)$, maximum of uncertainty). In conclusion, the following order chain holds for any probability distribution $\vec{p}$:

$$0 = H(\vec{p}_k) \le H(\vec{p}) \le H(\vec{p}_u) = \log N$$

where the value 0 corresponds to the minimum and $\log N$ to the maximum of uncertainty for any length N possible probability distribution $\vec{p}$.

2.1. **Measure distributions and probability distributions.** In this paper we are particularly interested to the so–called *measure distributions*, that is (non–trivial) vectors of the kind $\vec{m} = (m_1, m_2, \ldots, m_N)$ in which each component is non–negative ($\forall i,\ m_i \ge 0$), with at least an element $m_{i_0} \ne 0$, and such that their sum is not necessarily equal to 1: in general $\sum_{i=1}^N m_i = M(\vec{m})$ with $M(\vec{m}) \ne 0$ the *total measure* of the distribution, which depends from the particular measure distribution $\vec{m}$. For any measure distribution $\vec{m}$ it is possible to construct the corresponding probability distribution $\vec{p} = \left(\frac{m_1}{M(\vec{m})}, \frac{m_2}{M(\vec{m})}, \ldots, \frac{m_N}{M(\vec{m})}\right)$, which is the result of the normalization of the measure distribution $\vec{m}$ with respect to the quantity $M(\vec{m})$, i.e., $\vec{p} = \frac{1}{M(\vec{m})}\vec{m}$; for this reason a measure distribution is also called a *non–normalized probability distribution*. The entropy of $\vec{p}$, denoted by $H(\vec{m})$ instead of $H(\vec{p})$ in order to stress its dependence from the original measure distribution $\vec{m}$, is the sum of two terms

$$H(\vec{m}) = \log M(\vec{m}) - \frac{1}{M(\vec{m})} \sum_{i=1}^{N} m_i \log m_i \tag{7}$$

If one defines as *co-entropy* the quantity (also this depending from the measure distribution $\vec{m}$)

$$E(\vec{m}) = \frac{1}{M(\vec{m})} \sum_{i=1}^{N} m_i \log m_i \tag{8}$$

we have the following identity which holds for any *arbitrary* measure distribution:

$$H(\vec{m}) + E(\vec{m}) = \log M(\vec{m}) \tag{9}$$

The name co–entropy assigned to the quantity $E(\vec{m})$ rises from the fact that it "complements" the entropy $H(\vec{m})$ with respect to the value $\log M(\vec{m})$, which depends from the distribution $\vec{m}$. Of course, in the equivalence class of all measure distributions of identical total measure ($\vec{m}_1$ and $\vec{m}_2$ are equivalent iff $M(\vec{m}_1) = M(\vec{m}_2)$) this value is constant whatever be their length N.

3. Entropy and Co–Entropy of Partitions

We now want apply the above abstract results about entropy of generic probability distributions to the particular case of a *finite* universe X equipped with a partition $\pi = \{A_1, A_2, \ldots, A_N\}$, i.e., a collection of nonempty subsets A_i of X which are pairwise disjoints and whose set theoretic union is X. The subsets A_i from the partition π are the *elementary events* of the measure distribution $\vec{m}(\pi) = (|A_1|, |A_2|, \ldots, |A_N|)$, where the measure of the event A_i is the so–called *counting measure* $m_c(A_i) := |A_i|$. This measure distribution satisfies the conditions: (i) every $|A_i| > 0$; (ii) $\sum_{i=1}^N |A_i| = |X|$. The induced probability distribution is characterized by the following corresponding probabilities

$$p(A_i) = \frac{|A_i|}{|X|} \tag{10}$$

In this way we have generated the probability vector $\vec{p}(\pi) := (p(A_1), p(A_2), \ldots, p(A_N))$, depending from the partition π. The entropy of this probability distribution, simply written as $H(\pi)$ instead of $H(\vec{p}(\pi))$, is then

$$H(\pi) = -\sum p(A_i) \log p(A_i) = -\sum_{i=1}^{N} \frac{|A_i|}{|X|} \log \frac{|A_i|}{|X|} \tag{11}$$

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 chain of 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 $\pi = \pi_d$.

Remark 1. Recently, in the rough set community there is a certain attitude (see for instance [LS04]) to attribute to the Wierman paper [Wie99, sect. 4] the introduction of the entropy of a partition defined as (11). In [Wie99] this entropy has been called the *granularity measure*, and is considered as the quantity which “measures the uncertainty (in bits) associated with the prediction of outcomes where elements of each partition sets A_i are indistinguishable,” In [LS04] it is remarked that this “granularity measure” coincides with the Shannon entropy ([Sha48]), and it is interpreted as the “information measure of knowledge” furnished by the partition π.
To tell the truth, this entropy of partitions has been considered in the context of information theory several years before the Wiermann paper, see for instance the textbook [Rez94, p. 76] published in its first version in 1961. In this book one can found at p. 81 some other uniqueness characterizations of the entropy function of partitions (or measure of granularity in the Wiermann terminology) besides the one whose derivation can be found in [KW98], for instance the ones of D. A. Fadiev (1956) or of A. I. Khinchin (1953).

Note that the entropy (11) associated with a probability distribution π assumes also the following form:

$$H(\pi) = \log|X| - \frac{1}{|X|} \sum_{i=1}^{N} |A_i| \log|A_i| \tag{12}$$

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

$$E(\pi) := \frac{1}{|X|} \sum_{i=1}^{N} |A_i| \log|A_i| \tag{13}$$

then the (12) leads to the identity:

$$\forall \pi \in \Pi(X), \quad H(\pi) + E(\pi) = \log|X| \tag{14}$$

i.e., the quantity $E(\pi)$ is the “co–entropy” which *complements* the original entropy $H(\pi)$ with respect to the constant value $\log|X|$. This value, differently from the abstract case (9), is invariant with respect to the choice of the particular partition π of X.

Remark 2. 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 (13) has been considered in [LS04]. In the rough set context these authors called $E(\pi)$ with the name of *rough entropy of knowledge* π, whereas the standard Shannon notion of entropy $H(\pi)$ of (11) (or in the equivalent form (12)) has been called *granularity* or *information measure* of knowledge.
Let us note that this kind of rough entropy of knowledge has been introduced in [LX00] in the more general context of coverings some years before [LS04].

3.1. **Another measure distribution induced by partitions with associated pseudo co–entropy.** In order to better understand the application to coverings of the Liang–Xu (LX) approach to quantify information in the case of incomplete systems [LX00], let us now introduce (and compare with (13)) a new form of pseudo co–entropy related to a partition $\pi = \{A_1, \ldots, A_N\}$ by the following definition in which the sum involves the "local" information given by all the equivalence classes $\pi(x)$ for the point x ranging on the universe X:

$$E_{LX}(\pi) := \frac{1}{|X|} \sum_{x \in X} |\pi(x)| \cdot \log |\pi(x)| \tag{15}$$

Trivially, $\forall \pi \in \Pi(X)$, $0 = E_{LX}(\pi_d) \leq E_{LX}(\pi) \leq E_{LX}(\pi_t) = |X| \cdot \log |X|$. Moreover,

$$E_{LX}(\pi) = \frac{1}{|X|} \left[\sum_{x \in A_1} |\pi(x)| \cdot \log |\pi(x)| + \ldots + \sum_{x \in A_N} |\pi(x)| \cdot \log |\pi(x)| \right] \tag{16a}$$

$$= \frac{1}{|X|} \sum_{i=1}^{N} |A_i|^2 \cdot \log |A_i| \tag{16b}$$

and so from the fact that $1 \leq |A_i| \leq |A_i|^2$ it follows that $\forall \pi$, $0 \leq E(\pi) \leq E_{LX}(\pi)$. The comparison between (13) and (16b) put in evidence the very profound difference of these definitions. With the aim to capture some relationship with respect to a pseudo–entropy, for any partition π let us consider the vector

$$\vec{\mu}_\pi := \left(\mu_\pi(x) := \frac{|\pi(x)|}{|X|} : x \in X \right) \tag{17}$$

which is a *pseudo–probability distribution* since $\forall x$, $0 \leq \mu_\pi(x) \leq 1$, but $\mu(\pi) := \sum_{x \in X} \mu_\pi(x) = \sum_{i=1}^{N} \frac{|A_i|^2}{|X|} \geq 1$; this latter quantity is equal to 1 when $\sum_{i=1}^{N} |A_i|^2 = |X|$, and in this case the vector $\vec{\mu}_\pi$ defines a *probability distribution.* Moreover, for any partition it is $\mu(\pi) \leq |X|$, with $\mu(\pi_t) = |X|$. Applying in a pure formal way the formula (11) to this distribution one obtains

$$H_{LX}(\pi) = -\sum_{x \in X} \mu_\pi(x) \cdot \log \mu_\pi(x) = -\sum_{i=1}^{N} \frac{|A_i|^2}{|X|} \cdot \log \frac{|A_i|}{|X|}$$

from which it follows that:

$$H_{LX}(\pi) + E_{LX}(\pi) = \log |X| \cdot \mu(\pi)$$

Hence, $E_{LX}(\pi)$ is complementary to the "*pseudo–entropy*" $H_{LX}(\pi)$ with respect to the quantity $\log |X| \cdot \mu(\pi)$, which depends from the partition π by its "pseudo–measure" $\mu(\pi)$. For instance in the case of the trivial partition it is $H_{LX}(\pi_t) = 0$ and $E_{LX}(\pi_t) = |X| \cdot \log |X|$, with $H_{LX}(\pi_t) + E_{LX}(\pi_t) = |X| \cdot \log |X|$. On the other hand, in the case of the discrete partition it is $H_{LX}(\pi_d) = \log |X|$ and $E_{LX}(\pi_d) = 0$, with $H_{LX}(\pi_d) + E_{LX}(\pi_d) = \log |X|$.

Of course, the measure distribution (17) can be *normalized* by the quantity $\mu(\pi)$ obtaining a real probability distribution

$$\vec{\mu}^{(n)}(\pi) = \left(\mu_\pi^{(n)}(x) = \frac{|\pi(x)|}{\sum_{i=1}^{N} |A_i|^2} : x \in X \right)$$

But in this case the real entropy $H_{LX}^{(n)}(\pi) = -\sum_{x \in X} \mu_\pi^{(n)}(x) \cdot \log \mu_\pi^{(n)}(x)$ is linked to the above pseudo co–entropy (15) by the relationship $H_{LX}^{(n)}(\pi) + \frac{1}{\mu(\pi)} E_{LX}(\pi) = \log[\mu(\pi) \cdot |X|]$, in which the dependence from the partition π by the "measure" $\mu(\pi)$ is very hard to handle in applications.

3.2. **Anti–monotonicity of entropy and monotonicity of co–entropy with respect to the partition ordering.** On the family $\Pi(X)$ of all partitions of X let us introduce the following three binary relations:

(18a) $\pi_1 \preceq \pi_2$ iff $\forall A_i \in \pi_1,\ \exists B_j \in \pi_2 : A_i \subseteq B_j$

(18b) $\pi_1 \ll \pi_2$ iff $\forall B_j \in \pi_2,\ \exists \{A_{i_1}, A_{i_2}, \ldots, A_{i_p}\} \subseteq \pi_1 : B_j = A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_p}$

(18c) $\pi_1 \trianglelefteq \pi_2$ iff $\forall x \in X,\ \pi_1(x) \subseteq \pi_2(x)$

Note that these binary relations define the same binary relation on $\Pi(X)$ since trivially

$$\pi_1 \ll \pi_2 \quad \text{iff} \quad \pi_1 \preceq \pi_2 \quad \text{iff} \quad \pi_1 \trianglelefteq \pi_2$$

Remark 3. The introduction on $\Pi(X)$ of these indistinguishable binary relations $\preceq$, $\ll$, and $\trianglelefteq$ might seem a little bit redundant, but the reason of listing them in this partition context is essentially due to the fact that in the case of coverings of X they give rise to different relations as we will see in section (5). The structure $\langle \Pi(X), \preceq, \pi_d, \pi_t \rangle$ is a poset bounded by the *least* partition π_d and the *greatest* partition π_t: formally $\forall \pi \in \Pi(X),\ \ \pi_d \leq \pi \leq \pi_t$. This (finite) poset is a (complete) lattice with respect to the lattice meet $\pi_1 \wedge \pi_2 = \pi_1 \cdot \pi_2 := \{A_i \cap B_j : A_i \in \pi_1 \text{ and } B_j \in \pi_2\}$ and joint $\pi_1 \vee \pi_2 = \wedge\{\pi \in \Pi(X) : \pi_1 \preceq \pi \text{ and } \pi_2 \preceq \pi\}$.

Example 3.1. In the universe $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ let us consider the two partitions $\pi_1 = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7\}, \{8\}, \{9, 10\}\}$ and $\pi_2 = \{\{1, 2\}, \{3, 4\}, \{5, 6, 7\}, \{8, 9, 10\}\}$. Then their lattice meet is the new partition $\pi_1 \wedge \pi_2 = \{\{1, 2\}, \{3\}, \{4\}, \{5, 6\}, \{7\}, \{8\}, \{9, 10\}\}$ and the lattice join is the partition $\pi_1 \vee \pi_2 = \{\{1, 2, 3, 4, 5, 6, 7\}, \{8, 9, 10\}\}$.

If $\pi_1 \preceq \pi_2$ then (see also this example, in which $\pi_1 \wedge \pi_2 \preceq \pi_1, \pi_2 \preceq \pi_1 \vee \pi_2$) the partition π_1 (resp., π_2) is said to be *finer* (resp., *coarser*) than the partition π_2 (resp., π_1). If $\pi_1 \preceq \pi_2$ and $\pi_1 \neq \pi_2$ then π_1 (resp., π_2) is *strictly finer* (resp., *strictly coarser*) than π_2 (resp., π_1), written $\pi_1 \prec \pi_2$.

Remark 4. Given an information system $\langle X, Att, F \rangle$, the collection of all partitions generated by any possible family $\mathcal{A}$ of attributes $\{\pi(\mathcal{A}) \in \Pi(X) : \mathcal{A} \in \mathcal{P}(Att)\}$, owing to (3) is a sublattice of the lattice $\langle \Pi(X), \wedge, \vee, \pi_d, \pi_t \rangle$ of all partitions of X.
This sublattice in general does not contain the least element π_d unless one attribute $a_0 \in Att$ of the information system has the associated random variable $f_{a_0} : X \mapsto val(a_0)$ which is a bijection (equivalently, $|val(a_0)| = |X|$), and does not contain the greatest element π_t unless there exists an attribute a_1 whose set of values is a singleton: $val(a_1) = \{\alpha_1\}$.

Let us stress that a standard result of information theory assures the (strict) anti–monotonicity of entropy:

$$\pi_1 \prec \pi_2 \quad \text{implies} \quad H(\pi_2) < H(\pi_1) \tag{19}$$

Thus, the (19) and (14) lead to the following strict monotonicity of co–entropy with respect to the partition ordering (a direct proof of this result can be found in [LS04]):

$$\pi_1 \prec \pi_2 \quad \text{implies} \quad E(\pi_1) < E(\pi_2) \tag{20}$$

Remark 5. This result has been proved in the so–called *roughness monotonicity* theorem 10 of [Wie99], whose proof is based on lemma 8 of the same paper. In the partition context this result can be found for instance in [Rez94, eq. (3-36) at p. 84].

Let us now consider an information system $\langle X, Att, F \rangle$, where (as explained in section 1) any attribute $a \in Att$ is identified with a surjective function (random variable) $f_a : X \mapsto val(a)$ defined on X. Any family $\mathcal{A}$ of attributes from Att defines a partition $\pi(\mathcal{A})$ of the universe X according to the equivalence relation (4) and one can associate with this partition the entropy $H(\pi(\mathcal{A}))$ and the co–entropy $E(\pi(\mathcal{A}))$, according to (12) and (13) respectively. Each equivalence class of $\pi(\mathcal{A})$ furnishes then a *granule* of knowledge with respect to which any subset of X can be approximated. The partition π constitutes the collection of all *information granulation* which can be used on subsets of the universe.

Of course, if one increases the collection of attributes $\mathcal{A} \subset \mathcal{B}$, then with respect to the partial order of partitions one has that $\pi(\mathcal{B}) \preceq \pi(\mathcal{A})$ (with the possibility that $\pi(\mathcal{B}) = \pi(\mathcal{A})$), and from (19) and (20) it follows that

$$\mathcal{A} \subset \mathcal{B} \quad \text{implies} \quad H(\pi(\mathcal{A})) \leq H(\pi(\mathcal{B})) \tag{21a}$$

$$\mathcal{A} \subset \mathcal{B} \quad \text{implies} \quad E(\pi(\mathcal{B})) \leq E(\pi(\mathcal{A})) \tag{21b}$$

Hence, it is the entropy H (resp., E) which preserves the monotonicity (resp., anti–monotonicity) with respect to the set theoretic inclusion $\subseteq$ on the subsets of attributes.

Remark 6. If for two collections of attributes $\mathcal{A}$ and $\mathcal{B}$ the condition $\mathcal{A} \subset \mathcal{B}$ implies $\pi(\mathcal{A}) = \pi(\mathcal{B})$, then the family of attributes $\mathcal{B} \setminus \mathcal{A}$ is *redundant* with respect to the original knowledge furnished by $\mathcal{A}$. In this case the family *Att* can be, at least in line of principle, *reduct* eliminating all the attributes from $\mathcal{B}$ which are not in $\mathcal{A}$. Hence, we can say that an information system is *non–redundant* iff for any pairs of attribute collections $\mathcal{A}$ and $\mathcal{B}$, the condition $\mathcal{A} \subset \mathcal{B}$ implies $\pi(\mathcal{A}) \prec \pi(\mathcal{B})$.

3.3. **The knowledge granulation in partitions.** As we have seen in section 1, given a partition π of the universe X any subset Y of X can be approximated with respect to π by its lower and upper approximations (1), which in the present context we prefer to denote by $l_\pi(Y)$ and $u_\pi(Y)$ respectively. The rough approximation of Y, formalized by the pair $r_\pi(Y) = \langle l_\pi(Y), u_\pi(Y)\rangle$, describes the best approximation of Y from the bottom and the top making use of the granulation knowledge furnished by the equivalence classes from π as elementary sets or knowledge granules (events in the probability context).

In order to give a *measure* of the approximation $r_\pi(Y)$ of a generic subset Y of X, Pawlak introduced the two notions of *accuracy* and *roughness* (see for instance [Paw91]), formally defined respectively as follows:

$$\alpha_\pi(Y) = \frac{|l_\pi(Y)|}{|u_\pi(Y)|} \quad \text{and} \quad \rho_\pi(Y) = 1 - \alpha_\pi(Y) = \frac{|u_\pi(Y)| - |l_\pi(Y)|}{|u_\pi(Y)|} \tag{22}$$

With respect to this definition of roughness, the following monotonicity behavior holds:

$$\pi_1 \prec \pi_2 \quad \text{implies} \quad \forall Y \subseteq X : \rho_{\pi_1}(Y) \leq \rho_{\pi_2}(Y) \tag{23}$$

The two measures (22), each of which furnishes a numerical characterization of the rough approximation of Y with respect to π, have the withdraw of being unable to distinguish the roughness of the same set with respect to two different partitions. In other world, it could happen that a given subset Y has the same roughness measure $\rho_{\pi_1}(Y) = \rho_{\pi_2}(Y)$ also in the particular case of two partitions π_1, π_2 of which one is strictly finer than the other: $\pi_1 \prec \pi_2$

Example 3.2. In the universe $X = \{1, 2, 3, 4, 5, 6\}$, let us consider the two partitions $\pi_1 = \{\{1\}, \{2\},$
$\{3\}, \{4, 5, 6\}\}$ and $\pi_2 = \{\{1, 2\}, \{3\}, \{4, 5, 6\}\}$, with respect to which $\pi_1 \prec \pi_2$. The subset $Y = \{1, 2, 4, 6\}$ is such that $l_{\pi_1}(Y) = l_{\pi_2}(Y) = \{1, 2\}$ and $u_{\pi_1}(Y) = u_{\pi_2}(Y) = \{1, 2, 4, 5, 6\}$. This result implies that $\alpha_{\pi_1}(Y) = \alpha_{\pi_2}(Y)$ (or $\rho_{\pi_1}(Y) = \rho_{\pi_2}(Y)$).

A possible measure of roughness which is able to distinguish these unpleasant situations can be defined, according to [BPA98] (an extension of this definition to the covering case can be found in [LX00]), by the so–called *rough entropy* of Y *relative to the partition* π, which is the product of the Pawlak (local) roughness measure (22) of Y times the (global) co–entropy (13) of the partition π of X:

$$E_\pi(Y) = \rho_\pi(Y) \cdot E(\pi) \tag{24}$$

The *rough entropy* mapping $E_\pi : \mathcal{P}(X) \mapsto [0, \log |X|\,]$ can be considered a *local version of co–entropy* since it is "locally" defined on any subset Y of X, differently from the co–entropy (13) which is a global quantity assigned to the whole universe X. This local rough entropy (better, co–entropy) satisfies the following *strict monotonicity* condition, granted by the monotonicity of (23) and the strict monotonicity expressed in (20):

$$\pi_1 \prec \pi_2 \quad \text{implies} \quad \forall Y \subseteq X : \; E_{\pi_1}(Y) < E_{\pi_2}(Y)$$

In particular, for any subset Y one has that with respect to the discrete partition it is $E_{\pi_d}(Y) = 0$ and with respect to the trivial partition it is $E_{\pi_t}(Y) = \log|X|$, both independent from Y. And so, for any arbitrary partition π the following order chain holds: $0 = E_{\pi_d}(Y) \leq E_\pi(Y) \leq E_{\pi_t}(Y) = \log|X|$.

Now we want to discuss another approach to the local notion of rough entropy. Precisely, after the definition of the *exterior* of the subset Y as $e_\pi(Y) := X \setminus u_\pi(Y)$, one can introduce two new notions of accuracy and roughness as follows:

$$\hat{\alpha}_\pi(Y) = \frac{|l_\pi(Y)| + |e_\pi(Y)|}{|X|} \quad \text{and} \quad \hat{\rho}_\pi(Y){=}1 - \hat{\alpha}_\pi(Y) = \frac{|b_\pi(Y)|}{|X|}$$

With respect to these definitions, the following monotonicity behavior holds:

$$\pi_1 \prec \pi_2 \quad \text{implies} \quad \forall Y \subseteq X : \hat{\rho}_{\pi_1}(Y) \leq \hat{\rho}_{\pi_2}(Y) \tag{25}$$

The strict monotonicity in general does not hold, since notwithstanding $\pi_1 \prec \pi_2$ it could happen that for a certain Y the two boundaries are identical: $b_{\pi_1}(Y) = b_{\pi_2}(Y)$. This being stated, we can introduce the new *rough entropy* of Y relatively to the partition π by the following modification of (24):

$$\hat{E}_\pi(Y) = \hat{\rho}_\pi(Y) \cdot E(\pi) \tag{26}$$

From the monotonicity (25) and the strict monotonicity (20), the following preservation of strict monotonicity holds:

$$\pi_1 \prec \pi_2 \quad \text{implies} \quad \forall Y \subseteq X : \hat{E}_{\pi_1}(Y) < \hat{E}_{\pi_2}(Y) \tag{27}$$

In particular, for any subset Y one has that with respect to the discrete partition it is $\hat{E}_{\pi_d}(Y) = 0$ and with respect to the trivial partition it is $\hat{E}_{\pi_t}(Y) = \log|X|$, both independent from Y. And so, for any arbitrary partition $0 = \hat{E}_{\pi_d}(Y) \leq \hat{E}_\pi(Y) \leq \hat{E}_{\pi_t}(Y) = \log|X|$. Also in this case, (26) defines a *local notion* of co–entropy with respect to a given partition of the universe.

Let us remark that with respect to a given subset Y of the universe the granules of the partition π can be classified into three disjoint classes: $\pi_i(Y) := \{I \in \pi : I \subseteq Y\}$ (interior events), $\pi_e(Y) := \{E \in \pi : E \subseteq Y^c\}$ (exterior events), and $\pi_b(Y) := \pi \setminus (\pi_i(Y) \cup \pi_e(Y)) = \{B \in \pi : B \cap Y \neq \emptyset,\ B \cap Y^c \neq \emptyset\}$ (boundary events), moreover we introduce $\pi_u(Y) := \pi_i(Y) \cup \pi_b(Y)$ and $\pi_{i,e}(Y) := \pi_i(Y) \cup \pi_e(Y)$. With respect to this classification of π, the partition co–entropy (13) can be formalized as

$$E(\pi) = \frac{1}{|X|} \sum_{U \in \pi_{i,e}(Y)} |U| \log|U| + \frac{1}{|X|} \sum_{B \in \pi_b(Y)} |B| \log|B|$$

and so, setting $E^b_\pi(Y) = 1/|X| \sum_{B \in \pi_b(Y)} |B| \log|B|$ we can define the new *rough entropy* of Y relatively to the partition π:

$$\hat{E}_\pi(Y) = \hat{\rho}_\pi(Y) \cdot E^b_\pi(Y) \leq E_\pi(Y)$$

This new rough entropy locally defined for any subset Y gives a more precise measure of the roughness of Y with respect to the previous ones, since the following ordering chain holds:

$$0 \leq \hat{E}^b_\pi(Y) \leq \hat{E}_\pi(Y) \leq E_\pi(Y) \leq \log|X| \tag{28}$$

Moreover it satisfies the following monotonic correlation with respect to the partial order of partitions.

Proposition 3.1. *We have that:* $\pi_1 \prec \pi_2$ *implies* $\forall Y \subseteq X : \hat{E}_{\pi_1}(Y) \leq \hat{E}_{\pi_2}(Y)$.

Proof. From $\pi_1 \prec \pi_2$ it follows that for any $A \in (\pi_1)_b(Y)$ there exists at least one $B \in \pi_2$: $A \subseteq B$. This B is a fortiori an element of $A \in (\pi_2)_b(Y)$. Moreover, any of such B, in general, contains more than one element from $(\pi_1)_b(Y)$; let us denote by $A_1, A_2, \ldots, A_n$ these elements, then $A_1 \cup A_2 \cup \ldots \cup A_n \subseteq B$, and so $\sum_i |A_i| \leq |B|$. Then, we have that for every $i = 1, 2, \ldots, n$, $0 \leq \log|A_i| \leq \max\{\log|A_i|\}_i$, from which we get that $|A_i| \log|A_i| \leq |A_i|(\max\{\log|A_i|\}_i)$, and so $\sum_i |A_i| \log|A_i| \leq \sum_i |A_i|(\max\{\log|A_i|\}_i) \leq |B| \log B$. The thesis follows from this result and (25). □

This entropy of Y relatively to partitions in general is not strictly monotonic, since if under condition $\pi_1 \prec \pi_2$ it turns out that $(\pi_1)_b(Y) = (\pi_2)_b(Y)$ then both $\hat{\rho}_{\pi_1}(Y) = \hat{\rho}_{\pi_2}(Y)$ and $E^b_{\pi_1}(Y) = E^b_{\pi_2}(Y)$, leading to the equality of the corresponding rough entropies $\hat{E}_{\pi_1}(Y) = \hat{E}_{\pi_2}(Y)$.

Remark 7. Let us note that in any of the above local rough entropies the range of variation is between the value 0 and the value $\log|X|$. In particular, whatever be the involved entropy, the minimum is reached by the discrete partition π_d and the maximum by the trivial partition π_t. Furthermore, the discrete partition is characterized by the fact that the rough approximation of any subset H is the pair $r_d(H) = \langle H, H\rangle$, i.e., all the subsets of the universe are crisp, and the trivial partition by the property that $r_t(H) = \langle \emptyset, X\rangle$, i.e., it is invariant with respect to H and the unique crisp sets are the empty set and the whole universe.
Hence, one can adopt the closed interval $[0, \log|X|\,]$ as the *reference scale of crispness*, where the value 0 (resp., $\log|X|$) corresponds to the maximum (resp., minimum) of crispness, or minimum (resp., maximum) of roughness. From this point of view, and taking into account the (28), we can consider that the local entropy $\hat{E}^b_\pi(Y)$ furnishes the best valuation of the "degree' of crispness relatively to all the others.

4. Entropies and Co–Entropies of Coverings

In this section we generalize the notion of partition of a finite universe X introducing the following notion of covering.

Definition 4.1. Let X be a finite set (*universe*), then a *covering* of X is any finite collection $\gamma = \{B_1, B_2, \ldots, B_N\}$ of distinct subsets of X, also in this context called *elementary events*, such that (1) $B_i \neq \emptyset$ for all $B_i \in \gamma$ and (2) $\bigcup_{i=1}^N B_i = X$.

A partition is then a covering satisfying the further condition that its elements are pairwise disjoint. The collection of all coverings of the universe X will be denoted by $\Gamma(X)$. Of course, any partition is also a covering and so $\Pi(X) \subseteq \Gamma(X)$.

4.1. **Bridging the gap between partition and covering.**

In order to "bridge the gap" between partition and covering, let us note that, given a partition $\pi = \{A_1, A_2, \ldots, A_N\} \in \Pi(X)$, if one introduces the characteristic functional $\chi_{A_i} : X \mapsto \{0,1\}$ of the 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 to the partition π is a *crisp* (*sharp*) *identity resolution*, i.e., a family of crisp sets such that the following property holds:

$$\forall x \in X, \quad \sum_{i=1}^N \chi_{A_i}(x) = 1 \tag{29}$$

The *measure* of the elementary event A_i is defined as

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

with $\sum_{i=1}^N m(A_i) = |X|$ and so the probabilities (10) can also be expressed as

$$p(A_i) = \frac{1}{|X|} m(A_i) = \frac{1}{|X|} \sum_{x\in X} \chi_{A_i}(x) \tag{30a}$$

$$= \frac{1}{\sum_{i=1}^N m(A_i)}\, m(A_i) \tag{30b}$$

and the entropy (11) and co–entropy (13) associated with π respectively as

$$H(\pi) = \log|X| - \frac{1}{|X|}\sum_{i=1}^N m(A_i)\log m(A_i) \tag{31a}$$

$$E(\pi) = \frac{1}{|X|}\sum_{i=1}^N m(A_i)\log m(A_i) \tag{31b}$$

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

4.2. **Entropy of a covering: a first approach.** To any covering $\gamma = \{B_1, B_2, \ldots, B_N\}$ it is possible to associate the mapping $n : X \mapsto \mathbb{N}$ which counts the number of occurrences of the element x in γ according to the definition

$$\forall x \in X, \quad n(x) := \sum_{i=1}^{N} \chi_{B_i}(x) \tag{32}$$

Moreover, to any subset B_i of the covering γ one can introduce the corresponding *fuzzy set* $\omega_{B_1} : X \mapsto [0,1]$ defined as

$$\forall x \in X, \quad \omega_{B_i}(x) := \frac{1}{n(x)} \chi_{B_i}(x) \tag{33}$$

Remark 8. Let us note that in the particular case of a partition π of X, described by the crisp identity resolution $\mathcal{C}(\pi) = \{\chi_{A_1}, \chi_{A_2}, \ldots, \chi_{A_N}\}$ the number of occurrence of any point x expressed by (32), taking into account (29), is the identically 1 constant function $\forall x \in X$, $n(x) = 1$, and so the fuzzy set (33) is nothing else than the characteristic function itself: $\omega_{A_i} = \chi_{A_i}$.

This as a proof that the fuzzy sets of (33) are a generalization in the context of coverings of the representation of a partition by the identity resolution consisting of the corresponding characteristic functions. This is further one strengthened by the following result which assures that this fuzzy set representation of any covering is always an identity resolution.

Proposition 4.1. *The collection of fuzzy mappings $\mathcal{F}(\gamma) := \{\omega_{B_i} \in [0,1]^X : i = 1, 2, \ldots, N\}$ generated by a covering $\gamma = \{B_i : i = 1, 2, \ldots, N\}$ of the universe X according to* (33)*, is an* identity resolution*, in the sense that*

$$\forall x \in X, \quad \sum_{i=1}^{N} \omega_{B_i}(x) = 1$$

If one denotes by **1** *the identically 1 mapping ($\forall x \in X$, $\mathbf{1}(x) = 1$), then the previous identity resolution condition can be expressed as the functional identity $\sum_{i=1}^{N} \omega_{B_i} = \mathbf{1}$.*

Proof. Indeed, for every $x \in X$ one has $\sum_{i=1}^{N} \omega_{B_i}(x) = \frac{1}{n(x)} \sum_{i=1}^{N} \chi_{B_i}(x)$, and the thesis follows from (32). □

The measure of the generic "event" B_i of the covering γ is then defined as follows

$$m(B_i) := \sum_{x \in X} \omega_{B_i}(x) = \sum_{x \in X} \frac{1}{n(x)} \chi_{B_i}(x) \tag{34}$$

In this way we have defined a *covering measure* on the collection of all elementary events from the covering γ by the mapping $m : \{B_1, B_2, \ldots, B_N\} \mapsto \mathbb{R}_+$.

Proposition 4.2. *Let $\gamma = \{B_1, B_2, \ldots, B_N\}$ be a covering of the universe X, then the now introduced measure mapping satisfies the* normalization *condition:*

$$\sum_{i=1}^{N} m(B_i) = |X| \tag{35}$$

Moreover, for any elementary event B_i of the covering one has that:

$$0 \le m(B_i) \le |B_i| \le |X| \tag{36}$$

Proof.

$$\sum_{i=1}^{N} m(B_i) = \sum_{i=1}^{N}\Big(\sum_{x\in X}\omega_{B_i}(x)\Big) = \sum_{i=1}^{N}\Big(\sum_{x\in X}\frac{1}{n(x)}\chi_{B_i}(x)\Big)$$
$$= \sum_{x\in X}\frac{1}{n(x)}\Big(\sum_{i=1}^{N}\chi_{B_i}(x)\Big) = \sum_{x\in X}\frac{n(x)}{n(x)} = |X|$$

Moreover, from the relationship $1/n(x) \le 1$ true whatever be $x \in X$ it follows that

$$m(B_i) = \sum_{x\in X}\frac{1}{n(x)}\chi_{B_i}(x) \le \sum_{x\in X}\chi_{B_i}(x) = |B_i|$$

□

If one introduces the quantities $p(B_i) := \frac{1}{|X|}m(B_i)$, from (35) and (36) it follows that the vector $\vec{p}(\gamma) = (p(B_1), p(B_2), \ldots, p(B_N))$ defines a probability distribution induced by the covering γ. In particular, it results that

$$p(B_i) = \frac{1}{|X|}m(B_i) = \frac{1}{|X|}\sum_{x\in X}\omega_{B_i}(x) = \frac{1}{|X|}\sum_{x\in X}\frac{1}{n(x)}\chi_{B_i}(x) \tag{37}$$

The entropy of this probability distribution induced from the covering γ is then the real non–negative quantity

$$0 \le H(\gamma) = -\sum_{i=1}^{N} p(B_i)\log p(B_i) \le \log N \tag{38}$$

Trivially, by (37), one gets that

$$H(\gamma) = \log|X| - \frac{1}{|X|}\sum_{i=1}^{N} m(B_i)\log m(B_i) \tag{39}$$

and so also in this case we can introduce the *co–entropy* of the covering γ as the quantity

$$E(\gamma) = \frac{1}{|X|}\sum_{i=1}^{N} m(B_i)\log m(B_i) \tag{40}$$

obtaining from (39) the following identity

$$\forall\gamma \in \Gamma(X), \quad H(\gamma) + E(\gamma) = \log|X| \tag{41}$$

which is an extension to coverings of the identity (14) previously proved for any possible partition. Also in this case the "entropy" $E(\gamma)$ complements the original entropy $H(\gamma)$ with respect to the constant quantity $\log|X|$, invariant with respect to the choice of the covering γ. This co–entropy refers to the *measure distribution* (or *non normalized probability distribution*) $\vec{m}(\gamma) = \big(m(B_1), m(B_2), \ldots,$
$m(B_N)\big)$ for which the following hold: (1) every $m(B_i) \ge 0$; (2) $\sum_{i=1}^{N} m(B_i) = \log|X|$. Moreover, the corresponding (normalized) probability distribution $\vec{m}(\gamma)/|X| = \big(m(B_1)/|X|, m(B_2)/|X|, \ldots, m(B_N)/|X|\big)$, by the (37), coincides with the original probability distribution $\vec{p}(\gamma) = (p(B_1), p(B_2), \ldots, p(B_N))$.

Let us note that differently from the partition case, the now introduced co–entropy of a covering might have negative terms of the sum, precisely when $m(B_i) < 1$. And thus, the withdraw of this co–entropy is that it could assumes negative values.

Example 4.1. In the universe $X = \{1, 2, 3\}$, let us consider the covering $\gamma = \{\{1\}, \{1, 2\}, \{2, 3\}, \{2\}\}$. Then, $E(\gamma) \cong -0.2314$ and $H \cong 1,8163$, and so $E(\gamma) + H(\gamma) \cong 1.5850 \cong \log 3$.

Finally, the difference between partitions and coverings lies in the different definition of the measure m according to

$$\forall A_i \in \pi : m(A_i) = \sum_{x \in X} \chi_{A_i}(x) = |A_i| \qquad \text{and} \qquad \forall B_i \in \gamma : m(B_i) = \sum_{x \in X} \frac{1}{n(x)} \chi_{B_i}(x)$$

Remark 9. Note that in the measure of elementary events B_i from a covering γ a decisive role is played by the number $n(x)$ for x running in X.
Of course, since any partition π is also a covering one can apply the (39) to π but the result coincide with the (11), i.e., the just introduced covering entropy is a generalization of the standard notion of partition entropy.

4.3. **A second entropy of coverings inspired by the Liang and Xu approach.** An interesting approach to entropy of the covering induced by an *incomplete* information system has been recently introduced by Liang and Xu (LX) in [LX00]. Before entering into discussion of the LX approach, but inspired by it, we introduce in this subsection a second definition of covering entropy (the original LX approach to entropy applied to the similarity covering from any incomplete information system will be treated in the sequel). Precisely, an approach to entropy of coverings different from the one discussed in section 4.2 (also in this case based on another possible generalization to coverings of what happens in the case of partitions) can be obtained taking into account the counting measure $m_c(Y) = |Y|$ for any subset Y of X. Let us now consider a covering $\gamma = \{B_1, B_2, \dots, B_N\}$ of the universe X. The *total outer measure* of X induced from γ is then defined as the non–negative quantity

$$m^*(\gamma) := \sum_{i=1}^{N} |B_i| \geq |X| = m_c(X) \tag{42}$$

and a new probability of occurrence of the elementary event B_i from the covering γ is introduced according to

$$p^*(B_i) := \frac{|B_i|}{m^*(\gamma)} \tag{43}$$

The vector $\vec{p^*}(\gamma) := \big(p^*(B_1), p^*(B_2), \dots, p^*(B_N)\big)$ is a probability distribution since trivially: (1) every $p^*(B_i) \geq 0$; (2) $\sum_{i=1}^{N} p^*(B_i) = 1$. Then the entropy (6) in the case of this probability distribution is

$$H^*(\gamma) = \log m^*(\gamma) - \frac{1}{m^*(\gamma)} \sum_{i=1}^{N} |B_i| \log |B_i| \tag{44}$$

and so, introducing the *co–entropy*

$$E^*(\gamma) := \frac{1}{m^*(\gamma)} \sum_{i=1}^{N} |B_i| \log |B_i| \tag{45}$$

the above equation (44) leads to the identity

$$\forall \gamma \in \Gamma(X), \quad H^*(\gamma) + E^*(\gamma) = \log m^*(\gamma) \tag{46}$$

i.e., whatever be the covering γ of X the "entropy" $E^*(\gamma)$ is *complementary* to the original entropy $H^*(\gamma)$ with respect to the quantity $\log m^*(\gamma)$, which in this approach depends from the choice of the covering γ. The main difference of the co–entropy (45) with respect to the previous co–entropy (40) is that now $E^*(\gamma) \geq 0$ for every covering γ.

From the fact that $|X| \leq m^*(\gamma)$ and the monotonicity of the *log* function, by the comparison of (46) with (41) we obtain that

$$\forall \gamma \in \Gamma(X), \quad H(\gamma) + E(\gamma) \leq H^*(\gamma) + E^*(\gamma) \tag{47}$$

with the possibility of $E(\gamma) < 0$, whereas we stress that $\forall \gamma$, $H(\gamma) \geq 0$.

Remark 10. Also in this case one can apply all the obtained results to the case of partitions (which are particular coverings). The total outer measure (42) of a partition π is just the cardinality of the universe: $\forall \pi \in \Pi(X)$, $m^*(\pi) = |X|$.
Thus, in the case of a partition the entropy (44) coincides with the standard partition entropy of equation (11): $\forall \pi \in \Pi(X)$, $H(\pi) = H^*(\pi)$, obtaining also in this case that the new covering entropy is a generalization of the partition entropy.

4.4. **A third approach to entropy of coverings: the "global" LX–like.** If instead of taking into account the probability (43), but always inspired by [LX00], one considers the new probability of the elementary event B_i defined as

$$p_{LX}(B_i) := \frac{|B_i|}{|X|}$$

then the probability vector $\vec{p}_{LX}(\gamma) := (p_{LX}(B_1), p_{LX}(B_2), \ldots, p_{LX}(B_N))$ does *not* define a probability distribution also if for any event B_i, $0 \le p_{LX}(B_i) \le 1$, since in general $\sum_{i=1}^N p_{LX}(B_i) \ge 1$. Notwithstanding this withdraw one can try to define, in analogy with (11), the *pseudo–entropy* in the usual way

$$H_{LX}^{(g)}(\gamma) := -\sum_{i=1}^{N} p_{LX}(B_i) \log p_{LX}(B_i) = m^*(\gamma) \frac{\log |X|}{|X|} - \frac{1}{|X|} \sum_{i=1}^{N} |B_i| \log |B_i|$$

Thus, introduced the quantity

$$E_{LX}^{(g)}(\gamma) := \frac{1}{|X|} \sum_{i=1}^{N} |B_i| \log |B_i| \tag{48}$$

we obtain the following identity

$$\forall \gamma \in \Gamma(X), \quad H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma) = m^*(\gamma) \frac{\log |X|}{|X|} \tag{49}$$

and so the quantity $E_{LX}^{(g)}(\gamma)$ is complementary to the pseudo–entropy $H_{LX}^{(g)}(\gamma)$ (in other words it is a *pseudo co–entropy*) with respect to the quantity $m^*(\gamma)\frac{\log|X|}{|X|}$, which depends from the covering γ by the factor $m^*(\gamma)$. Since $\log |X| \le m^*(\gamma) \log |X|/|X|$, then we get

$$\forall \gamma \in \Gamma(X), \quad H(\gamma) + E(\gamma) \le H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma) \tag{50}$$

Summarizing, from (47) and (50) we have that

$$\forall \gamma \in \Gamma(X), \quad H(\gamma) + E(\gamma) \le \min\{H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma), H^*(\gamma) + E^*(\gamma)\}$$

The order relationship between $H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma)$ and $H^*(\gamma) + E^*(\gamma)$ depends obviously from the order relationship between $\log |X|/|X|$ and $\log m^*(\gamma)/m^*(\gamma)$, where the former is constant and the latter depends from the involved covering. Since the function $f(x) = \log x/x$ is monotonically decreasing for $x \ge 2$, if the cardinality of the universe is not trivial in the sense that $|X| \ge 2$, from the condition $|X| \le m^*(\gamma)$ it follows that $\log m^*(\gamma)/m^*(\gamma) \le \log |X|/|X|$, i.e., $\log m^*(\gamma) \le m^*(\gamma) \log |X|/|X|$. Then, from (46) and (49) we obtain

$$\text{If } |X| \ge 2, \text{ then } \forall \gamma \in \Gamma(X), \quad H^*(\gamma) + E^*(\gamma) \le H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma) \tag{51}$$

and so (50) and (51) lead to: If $|X| \ge 2$, then $\forall \gamma \in \Gamma(X)$, $\max\{H(\gamma) + E(\gamma), H^*(\gamma) + E^*(\gamma)\} \le H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma)$. We can conclude that under the condition of a sufficiently rich universe the following order chain holds:

$$\text{If } |X| \ge 2, \text{ then } \forall \gamma \in \Gamma(X),$$
$$H(\gamma) + E(\gamma) \le H^*(\gamma) + E^*(\gamma) \le H_{LX}^{(g)}(\gamma) + E_{LX}^{(g)}(\gamma)$$

As to the behavior of the co–entropies (45) and (48), trivially $E^*(\gamma) = \frac{|X|}{m^*(\gamma)} E_{LX}^{(g)}(\gamma)$, and so from the fact that $0 \le |X|/m^*(\gamma) \le 1$ we get that $\forall \gamma \in \Gamma(X)$, $E^*(\gamma) \le E_{LX}^{(g)}(\gamma)$.
On the other hand, comparing (40) with (45), from the fact that the function $h(p) = p \log p$

is negative for $0 \le p < 1$ and positive and monotonically increasing for $p > 1$, we get that if $m(B_i) \le 1$ then $m(B_i) \log m(B_i) \le 0 \le |B_i| \log |B_i|$ (recalling that $|B_i| \ge 1$). Moreover, if $m(B_i) > 1$ then from (36) it follows that also in this case $m(B_i) \log m(B_i) \le |B_i| \log |B_i|$. Thus, we can state that $\frac{1}{m^*(\gamma)} \sum_{i=1}^{N} m(B_i) \log m(B_i) \le \frac{1}{m^*(\gamma)} \sum_{i=1}^{N} |B_i| \log |B_i|$. From this result we get $\forall \gamma \in \Gamma(X)$, $\frac{|X|}{m^*(\gamma)} E(\gamma) \le E^*(\gamma)$. Summarizing, we have obtained the following order chain involving the previously introduced co–entropies

$$\forall \gamma \in \Gamma(X), \quad \frac{|X|}{m^*(\gamma)} E(\gamma) \le E^*(\gamma) \le E_{LX}^{(g)}(\gamma)$$

Taking into account that for any elementary event it is $|B_i| \le |X|$, and so $m^*(\gamma) = \sum_{i=1}^{N(\gamma)} |B_i| \le N(\gamma) \cdot |X| \le |\mathcal{P}(X)| \cdot |X|$ (with $|\mathcal{P}(X)|$ the cardinality of the power set $\mathcal{P}(X)$ of the universe X), the above chain of inequalities can be further on specialized by

$$\forall \gamma \in \Gamma(X), \quad \frac{1}{|\mathcal{P}(X)|} E(\gamma) \le E^*(\gamma) \le E_{LX}^{(g)}(\gamma)$$

Similarly, from (44) we get that $(m^*(\gamma)/|X|)\, H^*(\gamma) = \frac{m^*(\gamma)}{|X|} \log m^*(\gamma) - \frac{1}{|X|} \sum |B_i| \log |B_i| \ge \frac{m^*(\gamma)}{|X|} \log |X| - \frac{1}{|X|} \sum |B_i| \log |B_i|$ and so

$$\forall \gamma \in \Gamma(X), \quad H_{LX}^{(g)}(\gamma) \le \frac{m^*(\gamma)}{|X|} H^*(\gamma) \le |\mathcal{P}(X)| \cdot H^*(\gamma)$$

5. Partial quasi–order relations on coverings

Generalizing the standard binary relation on the set of all partitions in the form (18a), we introduce the following definition about coverings.

Definition 5.1. On the set of all coverings $\Gamma(X)$ of the universe X the following binary relation:

$$\gamma \preceq \delta \quad \text{iff} \quad \forall C_i \in \gamma,\ \exists D_j \in \delta : C_i \subseteq D_j \tag{52}$$

is a quasi–order relation, i.e., a reflexive and transitive, but in general non anti–symmetric relation [Bir67, p. 20]. In this case, we will say that γ is *finer* than δ or that δ is *coarser* than γ. The corresponding *strict* quasi–order relation is $\gamma \prec \delta$ iff $\gamma \preceq \delta$ and $\gamma \neq \delta$.

Trivially, the discrete partition π_d and the trivial partition π_t are such that for every other covering γ of X one has that $\pi_d \preceq \gamma \preceq \pi_t$. In general, the two conditions $\gamma \preceq \delta$ and $\delta \preceq \gamma$ do not imply that $\gamma = \delta$. When this happen we denote this fact by $\gamma \simeq \delta$; this is of course an equivalence relation and we will say that the two coverings are *equivalent*.

Remark 11. A covering is said to be *genuine* iff it does not contain redundant elements, formally iff condition $C_i \subseteq C_j$ implies $C_i = C_j$.
In the class of all genuine coverings $\Gamma_g(X)$ the binary relation $\preceq$ is an ordering. Indeed, let γ, δ be two genuine coverings of X such that $\gamma \preceq \delta$ and $\delta \preceq \gamma$. Then for $\forall C \in \gamma$, and using $\gamma \preceq \delta$, we have that $\exists D \in \delta$: $C \subseteq D$; but from $\delta \preceq \gamma$ it follows that there is also a $C' \in \gamma$ such that $D \in C'$, and so $C \subseteq D \subseteq C'$ and by the genuine condition of γ necessarily $C = D$. Vice versa, for every $D \in \delta$ there exists $C \in \gamma$ such that $D = C$.

The extension to the case of coverings of the ordering on partitions (18b), leads to a quasi-order relation on $\Gamma(X)$ which is different from the above (52). This quasi ordering is formalized by the following.

Definition 5.2. On the set of all coverings $\Gamma(X)$ of the universe X the following binary relation defines a quasi–ordering:

$$\gamma \ll \delta \quad \text{iff} \quad \forall D \in \delta,\ \exists \{C_1, C_2, \ldots, C_p\} \subseteq \gamma : D = C_1 \cup C_2 \cup \ldots \cup C_p$$

Also in this case $\pi_d \ll \gamma \ll \pi_t$ whatever be the covering γ.

In general it is not possible to state any relationship between the now introduced two quasi–orderings.

Example 5.1. In the universe $X = \{1,2,3,4,5,6,7,8\}$, let us consider the two coverings $\gamma = \{\{1,2,5\},\{3,4\},\{5,6\},\{6,7,8\}\}$ and $\delta = \{\{1,2,3,4,5\},\{3,4,6,7,8\}\}$. Trivially, $\gamma \ll \delta$, but with respect to $\{5,6\} \in \gamma$ there is no $D \in \delta$ such that $\{5,6\} \subseteq D$, i.e., $\gamma \not\preceq \delta$.

On the other hand, in the universe $X' = \{1,2,3,4\}$ let us consider the two coverings $\gamma' = \{\{1\},\{2,3\},\{3,4\}\}$ and $\delta' = \{\{1,2\},\{2,3,4\}\}$. Then, $\gamma' \preceq \delta'$, but with respect to $\{1,2\} \in \delta'$ there is no family $\{C_{i_1},\ldots,C_{i_p}\} \subseteq \gamma'$ such that $\{1,2\} = C_{i_1} \cup \ldots \cup C_{i_p}$, i.e., $\gamma' \not\ll \delta'$.

5.1. **Application to the similarity relation induced by a covering.** For any covering $\gamma = (C_1, C_2, \ldots, C_N)$ of the universe X,

(M1) let us introduce the binary relation of γ–*similarity* $\mathcal{S}_\gamma$ defined as

$$(x,y) \in \mathcal{S}_\gamma \quad \text{iff} \quad \exists C \in \gamma : \ x, y \in C$$

This relation is reflexive and symmetric, but in general is not transitive;

(M2) for any element $x \in X$ let us define the corresponding *similarity class* generated by x as follows:

$$\gamma_u(x) = \{y \in X : (y,x) \in \mathcal{S}_\gamma\} = \cup\{C \in \gamma : x \in C\}$$

(M3) the collection $\gamma_u = \{\gamma_u(x) : x \in X\}$ of all such similarity classes is a covering of X, called the *upper covering* generated by γ, for which both $\gamma \preceq \gamma_u$ and $\gamma \ll \gamma_u$ hold.

Extending to the case of coverings the binary relation (18c) introduced on the set of partitions, we can introduce a third quasi–order relation on $\Gamma(X)$ as follows:

$$\gamma \trianglelefteq_u \delta \quad \text{iff} \quad \forall x \in X,\ \gamma_u(x) \subseteq \delta_u(x) \tag{53}$$

Also this binary relation is not antisymmetric.

Example 5.2. In the universe $X = \{1,2,3,4,5,6\}$ let us consider the two (genuine) coverings $\gamma = \{\{1,2,3,5,6\},\ \{2,4,5,6\}\}$ and $\delta = \{\{1,2,3,5,6\},\{4,5,6\},\{2,4,6\}\}$. Trivially, $\delta \preceq \gamma$, but $\gamma \not\preceq \delta$.

Then $\gamma_u(1) = \gamma_u(3) = \{1,2,3,5,6\}$, $\gamma_u(2) = \gamma_u(5) = \gamma_u(6) = X$, $\gamma_u(4) = \{2,4,5,6\}$ and $\delta_u(1) = \delta_u(3) = \{1,2,3,5,6\}$, $\delta_u(2) = \delta_u(5) = \delta_u(6) = X$, $\delta_u(4) = \{2,4,5,6\}$. Thus, $\gamma \trianglelefteq_u \delta$ and $\delta \trianglelefteq_u \gamma$, but $\gamma \neq \delta$.

In a dual way, given a covering γ of the universe X, for every point $x \in X$ it is possible to introduce the subset $\gamma_l(x) := \cap\{C \in \gamma : x \in C\}$. And so the collection $\gamma_l := \{\gamma_l(x) : x \in X\}$ is another covering of X, called the *lower covering* generated by γ, such that $\gamma_l \preceq \gamma$ and $\gamma_l \ll \gamma$. Summarizing, and owing to the transitivity of the involved quasi–orderings, we can state that whatever be the covering γ of X one has that

$$\gamma_l \preceq \gamma \preceq \gamma_u \qquad \text{and} \qquad \gamma_l \ll \gamma \ll \gamma_u \tag{54}$$

Moreover, it is possible to assign to any point $x \in X$ the *granular rough approximation of x induced by γ* as the pair $r_\gamma(x) := \langle \gamma_l(x), \gamma_u(x) \rangle$. The set $\gamma_l(x)$ (resp., $\gamma_u(x)$) is the *lower or inner* (resp., *upper or outer*) x–granule generated by γ. Note that

$$\forall x \in X, \quad \gamma_l(x) \subseteq \gamma_u(x)$$

Similarly to (53) we define the following quasi–order relation on $\Gamma(X)$ as:

$$\gamma \trianglelefteq_l \delta \quad \text{iff} \quad \forall x \in X,\ \gamma_l(x) \subseteq \delta_l(x) \tag{55}$$

Thus, we introduce a stronger quasi–order relation on $\Gamma(X)$ which satisfies both (53) and (55), as follows:

$$\gamma \trianglelefteq \delta \qquad \text{iff} \qquad \gamma \trianglelefteq_l \delta \quad \text{and} \quad \gamma \trianglelefteq_u \delta \tag{56}$$

Example 5.3. Let us consider the example 5.2, where we have shown that $\delta \preceq \gamma$ and $\delta \trianglelefteq_u \gamma$. Moreover, $\gamma_l(1) = \gamma_l(3) = \{1,2,3,5,6\}$, $\gamma_l(2) = \gamma_l(5) = \gamma_l(6) = \{2,5,6\}$, $\gamma_l(4) = \{2,4,5,6\}$ and $\delta_l(1) = \delta_l(3) = \{1,2,3,5,6\}$, $\delta_l(2) = \{2,6\}$, $\delta_l(4) = \{4,6\}$, $\delta_l(5) = \{5,6\}$, $\delta_l(6) = \{6\}$. It follows that also $\delta \trianglelefteq_l \gamma$, and thus $\delta \trianglelefteq \gamma$.

Let us observe that $\gamma \preceq \delta$ implies $\gamma \trianglelefteq_u \delta$, but in general it does not imply that also $\gamma \trianglelefteq_l \delta$.

Example 5.4. In the universe $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, let us consider the following two (genuine) coverings $\gamma_1 = \{B_1 = \{1, 2, 3, 4, 5, 6\},\ B_2 = \{4, 5, 6, 7, 8, 9, 10\}\}$ and $\gamma_2 = \{A_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9\},\ A_2 = \{4, 5, 6, 7, 8, 9, 10\}\}$. Also in this case, we have that $\gamma_1 \preceq \gamma_2$, and that γ_1 and γ_2 are not comparable with respect to the quasi–ordering $\ll$ on coverings of X: $\gamma_1 \not\ll \gamma_2$ and $\gamma_2 \not\ll \gamma_1$. Moreover, $u_{\gamma_1}(1) = u_{\gamma_1}(2) = u_{\gamma_1}(3) = B_1$, $u_{\gamma_1}(4) = u_{\gamma_1}(5) = u_{\gamma_1}(6) = X$, $u_{\gamma_1}(j) = B_2$ for all $j = 7, 8, 9, 10$ and $u_{\gamma_2}(1) = u_{\gamma_2}(2) = u_{\gamma_2}(3) = A_1$, $u_{\gamma_2}(10) = A_2$, $u_{\gamma_2}(i) = X$ for all others $i \neq 1, 2, 3, 10$, and thus $\gamma_1 \trianglelefteq_u \gamma_2$

Let us now consider the lower granules: we have $l_{\gamma_1}(1) = l_{\gamma_1}(2) = l_{\gamma_1}(3) = B_1$, $l_{\gamma_1}(4) = l_{\gamma_1}(5) = l_{\gamma_1}(6) = \{4, 5, 6\}$, $l_{\gamma_1}(j) = B_2$ for all $j = 7, 8, 9, 10$ and $l_{\gamma_2}(1) = l_{\gamma_2}(2) = l_{\gamma_2}(3) = A_1$, $l_{\gamma_2}(10) = A_2$, $l_{\gamma_2}(i) = \{4, 5, 6, 7, 8, 9\}$ for all others $i \neq 1, 2, 3, 10$. In this case we have that $\gamma_1 \not\trianglelefteq_l \gamma_2$ since, for instance, $l_{\gamma_1}(7) = l_{\gamma_1}(8) = l_{\gamma_1}(9) = B_2 \not\subseteq l_{\gamma_2}(7) = l_{\gamma_2}(8) = l_{\gamma_2}(9) = \{4, 5, 6, 7, 8, 9\}$ and also $\gamma_2 \not\trianglelefteq_l \gamma_1$ since, for instance, $l_{\gamma_2}(1) = A_1 \not\subseteq B_1 = l_{\gamma_1}(1)$.

$$\text{If } \gamma \trianglelefteq_u \delta \text{ (with } \gamma \not\preceq \delta\text{), then in general it is not true that also } \gamma \trianglelefteq_l \delta.$$

Example 5.5. In example 5.2 we have directly verified that $\gamma \trianglelefteq_u \delta$ (with $\gamma \not\preceq \delta$). But in this case $\gamma \not\trianglelefteq_l \delta$, since for instance $\gamma_l(6) = \{2, 5, 6\} \not\subseteq \{6\} = \delta_l(6)$.

As a completion of the above covering rough approximations (54), it is now trivial to show that for every covering γ also

$$\gamma_l \trianglelefteq \gamma \trianglelefteq \gamma_u$$

Thus, once constructed the two families of coverings $\Gamma_l(X) := \{\gamma_l : \gamma \in \Gamma(X)\}$ and $\Gamma_u(X) := \{\gamma_u : \gamma \in \Gamma(X)\}$, the pair $r(\gamma) := \langle \gamma_l, \gamma_u \rangle$ is the *rough approximation of the covering* γ with respect to all quasi–orderings $\preceq$, $\ll$, and $\trianglelefteq$ in the *rough approximation space* (see [Cat98, CC06]) $\langle \Gamma(X), \Gamma_l(X), \Gamma_u(X) \rangle$ consisting of the collection $\Gamma(X)$ of all *approximable coverings*, and the collections $\Gamma_l(X)$ and $\Gamma_u(X)$ of *inner* and *outer definable coverings*. Of course, any partition $\pi \in \Pi(X)$ is a *sharp covering* since its rough approximation $r(\pi) = \langle \pi_l, \pi_u \rangle$ satisfies the sharpness condition $\pi_l = \pi_u = \pi$.

Remark 12. On $\Gamma(X)$ it is possible to introduce a further quasi–order relation (and compare with equation (56)) according to

$$\gamma \Subset \delta \quad \text{iff} \quad \delta \trianglelefteq_l \gamma \ \text{ and } \ \gamma \trianglelefteq_u \delta \tag{57}$$

The following holds:

$$\gamma \Subset \delta \quad \text{implies} \quad \forall x \in X,\ \ \delta_l(x) \subseteq \gamma_l(x) \subseteq (???) \subseteq \gamma_u(x) \subseteq \delta_u(x) \tag{58}$$

where the question marks mean some undefinable, an so hidden, covering granule $\gamma(x)$ intermediate between the corresponding lower and upper granules. In this way we have the "local" behavior:

$$\forall x \in X,\ \ r_\gamma(x) := \langle \gamma_l(x), \gamma_u(x) \rangle \sqsubseteq \langle \delta_l(x), \delta_u(x) \rangle =: r_\delta(x)$$

Where $\sqsubseteq$ means that for any point x the local approximation $r_\gamma(x)$ given by the covering γ is better than the local approximation $r_\delta(x)$ given by the covering δ, according to (58). This behavior can be formalized by the compact global notation $r(\gamma) \sqsubseteq r(\delta)$ and so we can summarize

$$\gamma \Subset \delta \quad \text{implies} \quad r(\gamma) \sqsubseteq r(\delta)$$

5.2. **Some useful counter–examples of entropies and co–entropies behavior with respect to the quasi–orderings.** Unfortunately, the entropies (resp., co–entropies) expressed by (39) and (44) (resp., (40) and (45)) do not behave regularly, i.e., monotonically, with respect to the partial quasi–order relations among coverings expressed by (52) (the behavior with respect to the quasi–ordering (53) needs some more deep investigations, which are outside the scope of the present paper). The following two examples highlight this irregular behavior.

Example 5.6. In the universe $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ let us consider the (genuine) coverings treated in example 5.4 for which we recall that $\gamma_1 \preceq \gamma_2$, but γ_1 and γ_2 are not comparable with respect to both the quasi–orderings $\ll$ and $\trianglelefteq_l$. Then with respect to $\preceq$ we have the desired behavior of monotonicity for the co–entropies E, E^* and $E_{LX}^{(g)}$, and of anti–monotonicity for the entropies H, H^* and $H_{LX}^{(g)}$, as illustrated in table 1.

	E	H	E^*	H^*	$E_{LX}^{(g)}$	$H_{LX}^{(g)}$
γ_1	2.32915	0.99277	2.70471	0.99573	3.51613	0.80238
γ_2	2.35097	0.97095	3.01130	0.98870	4.81808	0.49700

TABLE 1. Entropies and co–entropies for γ_1 and γ_2, with $\gamma_1 \preceq \gamma_2$ and $\gamma_1 \trianglelefteq_u \gamma_2$, but $\gamma_1 \ntrianglelefteq_l \gamma_2$ and $\gamma_2 \ntrianglelefteq_l \gamma_1$; in particular $E(\gamma_1) < E(\gamma_2)$, $E^*(\gamma_1) < E^*(\gamma_2)$, and $E_{LX}^{(g)}(\gamma_1) < E_{LX}^{(g)}(\gamma_2)$.

Let us now consider in the same universe the two (genuine) coverings $\delta_1 = \{B_1 = \{1,2\},\ B_2 = \{2,3,4,5,6,7,8,9,10\}\}$ and $\delta_2 = \{A_1 = \{1,2,3\},\ A_2 = B_2 = \{2,3,4,5,6,7,8,9,10\}\}$.
Also in this case $\delta_1 \preceq \delta_2$, with δ_1 and δ_2 non–comparable with respect to the quasi–ordering $\ll$. Moreover, $u_{\delta_1}(1) = B_1$, $u_{\delta_1}(2) = X$, $u_{\delta_1}(j) = B_2$ for all others $j \neq 1,2$, and $u_{\delta_2}(1) = A_1$, $u_{\delta_2}(2) = u_{\delta_2}(3) = X$, $u_{\delta_2}(i) = A_2$ for all others $i \neq 1,2,3$; and thus $\delta_1 \trianglelefteq_u \delta_2$.
The corresponding lower granules in this case are: $l_{\delta_1}(1) = B_1$, $l_{\delta_1}(2) = 2$, $l_{\delta_1}(j) = B_2$ for all others $j \neq 1,2$, and $l_{\delta_2}(1) = A_1$, $l_{\delta_2}(2) = l_{\delta_2}(3) = \{2,3\}$, $l_{\delta_2}(i) = A_2$ for all others $i \neq 1,2,3$; and thus $\delta_1 \ntrianglelefteq_l \delta_2$. In fact we have that $l_{\delta_1}(3) = B_2 \nsubseteq l_{\delta_2}(3) = \{2,3\}$. Note that we also have $\delta_2 \ntrianglelefteq_l \delta_1$ since, for instance, $l_{\delta_2}(1) = A_1 \nsubseteq l_{\delta_1}(1) = B_1$.

As illustrated in table 2, in this case we have the undesired behavior of monotonicity for the entropies H and H^*, and of anti–monotonicity for the co–entropies E and E^* with respect to both $\preceq$ and $\trianglelefteq_u$. Whereas, as it happened in the previous case, also now the co–entropy $E_{LX}^{(g)}$ behaves as expected.

	E	H	E^*	H^*	$E_{LX}^{(g)}$	$H_{LX}^{(g)}$
δ_1	2.71209	0.60984	2.77539	0.68404	3.05293	0.60119
δ_2	2.60000	0.72193	2.77368	0.81128	3.32842	0.65789

TABLE 2. Entropies and co–entropies for δ_1 and δ_2, with $\delta_1 \preceq \delta_2$, $\delta_1 \trianglelefteq_u \delta_2$, but $\delta_1 \ntrianglelefteq_l \delta_2$ and $\delta_2 \ntrianglelefteq_l \delta_1$; in particular $E(\delta_1) > E(\delta_2)$ and $E^*(\delta_1) > E^*(\delta_2)$, but as expected $E_{LX}^{(g)}(\delta_1) < E_{LX}^{(g)}(\delta_2)$.

In conclusion of these examples, there is no expected monotonicity regularity of the involved entropies (or co–entropies) H and H^* (resp., E and E^*) with respect to the quasi–orderings $\preceq$ and $\trianglelefteq_u$ on coverings. This withdraw occurs also for the global LX co–entropy of (48),

Example 5.7. Making reference to the examples 5.2 and 5.3 we have the following table 3 involving the introduced global entropies and co–entropies:

	E	H	E^*	H^*	$E_{LX}^{(g)}$	$H_{LX}^{(g)}$
δ	1.14944	1.43552	1.91995	1.53948	3.51990	1.21920
γ	1.60509	0.97987	2.17885	0.99108	3.26827	0.60917

TABLE 3. Entropies and co–entropies for γ and δ, with $\delta \preceq \gamma$ and $\delta \trianglelefteq \gamma$, but $E_{LX}^{(g)}(\gamma) < E_{LX}^{(g)}(\delta)$.

6. LX entropies and co–entropies generated by a covering

We have seen that from any covering γ of the universe X it is possible to induce its lower γ_l and upper γ_u coverings consisting of the lower granules $\gamma_l(x)$ and upper granules $\gamma_u(x)$ for x ranging on the space X. In this way, besides the (globally defined) LX co–entropy (48) (and generalizing the procedure of subsection 3.1) it is possible to introduce two (locally defined) LX entropies (resp., co–entropies), named the lower and upper LX entropies (resp., co–entropies) respectively according to the following:

$$H_{LX}(\gamma_j) : = -\sum_{x\in X} \frac{|\gamma_j(x)|}{|X|} \log \frac{|\gamma_j(x)|}{|X|} \quad \text{for} \quad j = l, u \tag{59a}$$

$$E_{LX}(\gamma_j) : = \frac{1}{|X|} \sum_{x\in X} |\gamma_j(x)| \log |\gamma_j(x)| \quad \text{for} \quad j = l, u \tag{59b}$$

with the relationships

$$H_{LX}(\gamma_j) + E_{LX}(\gamma_j) = (\sum_{x\in X} |\gamma_j(x)|) \cdot \frac{\log |X|}{|X|}$$

The following holds.

Proposition 6.1. *Let γ_1 and γ_2 be two coverings of X such that $\gamma_1 \lhd_j \gamma_2$ for $j = l, u$, then $E_{LX}(\gamma_{1j}) < E_{LX}(\gamma_{2j})$. In particular, with respect to the quasi–ordering* (57) *we have that $\gamma_1 \Subset \gamma_2$ implies*

$$E_{LX}(\gamma_{2l}) \leq E_{LX}(\gamma_{1l}) \leq (???) \leq E_{LX}(\gamma_{1u}) \leq E_{LX}(\gamma_{2u}) \tag{60}$$

Proof. Trivial consequence of the fact that for every $x \in X$, $\gamma_{1j}(x) \subseteq \gamma_{2j}(x)$, with $1 \leq |\gamma_{1j}(x)| \leq |\gamma_{2j}(x)|$. So we are in the range $x \in [1,\infty)$ in which the mapping $x \log x$ is monotonically increasing, and from this we have that $E_{LX}(\gamma_{1j}) \leq E_{LX}(\gamma_{2j})$.

Moreover, $\gamma_1 \lhd \gamma_2$ means that there exists a $x_0 \in X$: $\gamma_{1j}(x_0) \subset \gamma_{2j}(x_0)$, i.e., such that $|\gamma_{1j}(x_0)| < |\gamma_{2j}(x_0)|$. From this fact it follows that $E_{LX}(\gamma_{1j}) < E_{LX}(\gamma_{2j})$. □

Remark 13. This result cannot be extended to the entropy (59a), since only in the range $x \in [0, 1/e]$ the quantity $x \log x$ is monotonically increasing, whereas in the other range $x \in [1/e, 1]$ it is monotonically decreasing.

Since for every point $x \in X$ the following set theoretic inclusions hold: $\gamma_l(x) \subseteq \gamma_u(x)$, with $1 \leq |\gamma_l(x)| \leq |\gamma_u(x)| \leq |X|$, it is possible to introduce the *rough co–entropy approximation* of the covering γ as the ordered pair of non–negative numbers:

$$r_E(\gamma) := \langle E_{LX}(\gamma_l),\, E_{LX}(\gamma_u)\rangle\,, \quad \text{with } 0 \leq E_{LX}(\gamma_l) \leq E_{LX}(\gamma_u) \leq |X| \cdot \log |X|$$

and so from (60) we have that

$$\gamma_1 \Subset \gamma_2 \quad \text{implies} \quad r_E(\gamma_1) \sqsubseteq r_E(\gamma_2)$$

where as usual for any 4–uple of non negative real numbers $r_l, r_u, p_l, p_u \in \mathbb{R}_+$, with $r_l \leq r_u$ and $p_l \leq p_u$, we denote by $\langle r_l, r_u\rangle \sqsubseteq \langle p_l, p_u\rangle$ the fact that $p_l \leq r_l \leq r_u \leq p_u$.

Example 6.1. In the case of the coverings of example 5.4 (and compared with the behavior of table 1) we have the results of the LX entropy (59a) co–entropy (59b) for γ_1 and γ_2 illustrated in table 4.

	E_{LX}	H_{LX}
$(\gamma_1)_l$	13.93999	4.33061
$(\gamma_2)_l$	19.82981	3.42369
$(\gamma_1)_u$	22.47931	2.76734
$(\gamma_2)_u$	30.45551	0.77061

TABLE 4. Entropies and co–entropies for γ_1 and γ_2 (with $\gamma_1 \preceq \gamma_2$ and $\gamma_1 \unlhd_u \gamma_2$, but $\gamma_1 \ntrianglelefteq_l \gamma_2$ and $\gamma_2 \ntrianglelefteq_l \gamma_1$); in particular $E_{LX}(\gamma_{1u}) \leq E_{LX}(\gamma_{2u})$.

Let us now consider the coverings of example 5.6. The LX entropy (59a) and co–entropy (59b) for δ_1 and δ_2 are illustrated in table 5 .

	E_{LX}	H_{LX}
$(\delta_1)_l$	23.02346	1.89100
$(\delta_2)_l$	20.84602	2.40784
$(\delta_1)_u$	25.45760	1.78221
$(\delta_2)_u$	27.08987	1.47871

TABLE 5. Entropies and co–entropies for δ_1 and δ_2 (with $\delta_1 \preceq \delta_2$ and $\delta_1 \trianglelefteq_u \delta_2$, but $\delta_1 \ntrianglelefteq_l \delta_2$ and $\delta_2 \ntrianglelefteq_l \delta_1$); in particular $E_{LX}(\delta_{1u}) \leq E_{LX}(\delta_{2u})$.

Example 6.2. Making reference to the examples 5.2 and 5.3 we have the following table 6:

	E_{LX}	H_{LX}
δ_l	4.86988	2.45418
γ_l	7.58066	2.32837
δ_u	12.95810	0.82837
γ_u	12.95810	0.82837

TABLE 6. Entropies and co–entropies for γ and δ, with $\gamma \preceq \delta$ and $\delta \trianglelefteq \gamma$; according to (60) we have that $E_{LX}(\delta_l) < E_{LX}(\gamma_l) < E_{LX}(\gamma_u) = E_{LX}(\delta_u)$.

6.1. **Application to incomplete information system.** As anticipated in section 1, given an incomplete information system $\mathcal{IS} := \langle X, Att, F\rangle$, for any family $\mathcal{A}$ of attributes it is possible to define on the objects of X the similarity relation $\mathcal{S}_\mathcal{A}$ given in equation (5). This relation generates a covering of the universe X through the *granules of information* (also *similarity classes*) $s_\mathcal{A}(x) = \{y \in X : (x, y) \in \mathcal{S}_\mathcal{A}\}$, since $X = \cup\{s_\mathcal{A}(x) : x \in X\}$ and $x \in s_\mathcal{A}(x) \neq \emptyset$. In the sequel this covering will be denoted by $\gamma(\mathcal{A}) := \{s_\mathcal{A}(x) : x \in X\}$ and their collection by $\Gamma(\mathcal{IS}) := \{\gamma(\mathcal{A}) \in \Gamma(X) : \mathcal{A} \subseteq Att\}$. With respect to this covering $\gamma(\mathcal{A})$, and in analogy with (59), we can introduce the two LX entropy and co–entropy as follows:

$$H_{LX}(\gamma(\mathcal{A})) = \sum_{x\in X} \frac{|s_\mathcal{A}(x)|}{|X|} \log \frac{|s_\mathcal{A}(x)|}{|X|} \quad \text{and} \quad E_{LX}(\gamma(\mathcal{A})) = \frac{1}{|X|} \sum_{x\in X} |s_\mathcal{A}(x)| \cdot \log |s_\mathcal{A}(x)| \tag{61}$$

Let us note that in general we do not obtain a *genuine covering*, i.e., a covering such that condition $H \subseteq K$ implies $H = K$ for arbitrary pair of its elements H, K.

Example 6.3. Let us consider the incomplete information system of table 7.

Flat	Price	Rooms	Down-Town	Furniture
f_1	high	2	yes	*
f_2	high	*	yes	no
f_3	*	2	yes	no
f_4	low	*	no	no
f_5	low	1	*	no
f_6	*	1	yes	*

TABLE 7. Flats incomplete information system.

If one considers the set of all attributes (i.e., $\mathcal{A} = Att(X)$) and the induced similarity relation according to (5) we have the following similarity classes (in which for the sake of simplicity we omit the denote the subscript $\mathcal{A}$):

$$s(f_1) = s(f_3) = \{f_1, f_2, f_3\}$$
$$s(f_2) = \{f_1, f_2, f_3, f_6\} \qquad s(f_4) = \{f_4, f_5\}$$
$$s(f_5) = \{f_4, f_5, f_6\} \qquad s(f_6) = \{f_2, f_5, f_6\}$$

This covering is not genuine since, for instance, $s(f_1) = s(f_3) \subset s(f_2)$.

Clearly, the collection $\gamma = \{s(f_j) : j = 1, \ldots, 6\}$ is a covering of X. Applying to this covering the considerations introduced in section 5.1, one has the following lower and upper granules generated by any point of the universe:

$$\gamma_l(f_1) = \gamma_l(f_3) = \{f_1, f_2, f_3\} \qquad \gamma_u(f_1) = \gamma_u(f_3)\{f_1, f_2, f_3, f_6\}$$
$$\gamma_l(f_2) = \{f_2\} \qquad \gamma_u(f_2) = \{f_1, f_2, f_3, f_5, f_6\}$$
$$\gamma_l(f_4) = \{f_4, f_5\} \qquad \gamma_u(f_4) = \{f_4, f_5, f_6\}$$
$$\gamma_l(f_5) = \{f_5\} \qquad \gamma_u(f_5) = \{f_2, f_4, f_5, f_6\}$$
$$\gamma_l(f_6) = \{f_6\} \qquad \gamma_u(f_6) = X$$

with the general behavior

$$\forall f_i \in X, \quad \gamma_l(f_i) \subseteq s(f_i) \subseteq \gamma_u(f_i) \tag{62}$$

Let us observe that lower granules constitute a covering, and the same happens with the upper granules. We can thus compute the corresponding entropies H_{LX} and co–entropies E_{LX}. We denote these two particular coverings respectively by γ_l and γ_u of X. From (62) we have that $\gamma_l \preceq \gamma \preceq \gamma_u$. In table 8 we can find the entropies and co–entropies (59) and (61) relative to γ, γ_l and γ_u of the present example.

	E_{LX}	H_{LX}
γ_l	1.91830	2.82080
γ	4.83659	2.91830
γ_u	9.31238	1.88912

TABLE 8. Entropies and co–entropies for γ, γ_l and γ_u induced by $\mathcal{A} = Att(X)$, with the non–monotonic behavior of the entropy H_{LX}.

Let us now consider the family of attributes $\mathcal{B} = \{Price, Rooms\}$. As in the previous case of $\mathcal{A} = Att(X)$, let us consider the induced similarity relation according to (5). Thus we obtain the covering $\delta = \{s_{\mathcal{B}}(f_j) : j = 1, \ldots, 6\}$ of X, constituted by the similarity classes:

$$s_{\mathcal{B}}(f_1) = \{f_1, f_2, f_3\} \qquad s_{\mathcal{B}}(f_2) = \{f_1, f_2, f_3, f_6\}$$
$$s_{\mathcal{B}}(f_3) = \{f_1, f_2, f_3, f_4\} \qquad s_{\mathcal{B}}(f_4) = \{f_3, f_4, f_5, f_6\}$$
$$s_{\mathcal{B}}(f_5) = \{f_4, f_5, f_6\} \qquad s_{\mathcal{B}}(f_6) = \{f_2, f_4, f_5, f_6\}$$

The lower and upper granules generated by any point of the universe are:

$$l_\delta(f_1) = \{f_1, f_2, f_3\} \qquad u_\delta(f_1) = \{f_1, f_2, f_3, f_4, f_6\}$$
$$l_\delta(f_2) = \{f_2\} \qquad u_\delta(f_2) = X$$
$$l_\delta(f_3) = \{f_3\} \qquad u_\delta(f_3) = X$$
$$l_\delta(f_4) = \{f_4\} \qquad u_\delta(f_4) = X$$
$$l_\delta(f_5) = \{f_4, f_5, f_6\} \qquad u_\delta(f_5) = \{f_2, f_3, f_4, f_5, f_6\}$$
$$l_\delta(f_6) = \{f_6\} \qquad u_\delta(f_6) = X$$

Let us observe that in this example we have that $\gamma \preceq \delta$, $\gamma \trianglelefteq_u \delta$, whereas we do not have $\gamma \trianglelefteq_l \delta$. The entropies and co–entropies (59) and (61) for δ_l, δ and δ_u are illustrated in table 9.

	E_{LX}	H_{LX}
δ_l	1.58496	2.72331
δ	6.91830	2.55990
δ_u	14.20973	0.43839

Table 9. Entropies and co–entropies for δ, δ_l and δ_u induced by $\mathcal{B} = \{Price, Rooms\}$.

Now, we want to make some consideration and give some results about the LX-approach to entropy in this context. Using the above definition of similarity classes, it is possible to define another order relation among coverings generated by all possible subfamilies of attributes as follows:

$$\gamma(\mathcal{A}) \leq_s \gamma(\mathcal{B}) \quad \text{iff} \quad \forall x \in X,\ s_{\mathcal{A}}(x) \subseteq s_{\mathcal{B}}(x) \tag{63}$$

Clearly, if $\gamma(\mathcal{A}) \leq_s \gamma(\mathcal{B})$ then $\gamma(\mathcal{A})$ (resp., $\gamma(\mathcal{B})$) is *finer* (resp., *coarser*) than $\gamma(\mathcal{B})$ (resp., $\gamma(\mathcal{A})$) inside $\Gamma(\mathcal{IS})$. In a similar way a strict order relation can also be defined as

$$\gamma(\mathcal{A}) <_s \gamma(\mathcal{B}) \quad \text{iff} \quad \forall x \in X,\ s_{\gamma(\mathcal{A})}(x) \subseteq s_{\gamma(\mathcal{B})}(x) \quad \text{and} \quad \exists x \in X : s_{\gamma(\mathcal{A})}(x) \neq s_{\gamma(\mathcal{B})}(x)$$

The following proposition states that there is a monotonic behavior between this order of coverings and the pseudo co–entropy of equation (48).

Proposition 6.2. [LX00] *Let $\mathcal{IS} := \langle X, Att, F\rangle$ be an incomplete information system and $\gamma(\mathcal{A})$ and $\gamma(\mathcal{B})$ two coverings generated by the subsets of attributes $\mathcal{A} \subseteq Att$ and $\mathcal{B} \subseteq Att$. If $\gamma(\mathcal{A}) \leq_s \gamma(\mathcal{B})$ (resp., $\gamma(\mathcal{A}) <_s \gamma(\mathcal{B})$), then $E_{LX}(\gamma(\mathcal{A})) \leq_s E_{LX}(\gamma(\mathcal{B}))$ (resp., $E_{LX}(\gamma(\mathcal{A})) <_s E_{LX}(\gamma(\mathcal{B}))$).*

Corollary 6.1. [LX00] *Let $\langle X, Att, F\rangle$ be an incomplete information system, and $\mathcal{A}$, $\mathcal{B}$ two subsets of attributes such that $\mathcal{A} \subseteq \mathcal{B}$. If $\gamma(\mathcal{A})$ (resp., $\gamma(\mathcal{B})$) is the covering relative to attributes $\mathcal{A}$ (resp., $\mathcal{B}$) then $E_{LX}(\gamma(\mathcal{A})) \leq_s E_{LX}(\gamma(\mathcal{B}))$.*

7. Conclusions

In this work we presented several ways to generalize to the case of coverings the notions of entropy and co-entropy usually defined for partitions. The aim is to be able to compare these notions to the case of different coverings induced by the same incomplete information system. Thus, some notions of quasi–order are introduced and it is proved that only the monotonicity of the LX "local" co–entropies (59b) with respect to the orderings $\trianglelefteq_j$ and $\Subset$ behave as expected. On the contrary, with a lot of counter–examples it is shown that all the involved covering entropies do not preserve the anti–monotonicity with respect to any of the introduced quasi–orderings on coverings. This asymmetric behavior leads to the conclusion that, differently from the case of partitions, it is the notion of co–entropy which plays an important role with respect to a quantification of a local notion of roughness in the case of coverings.

On this issue, some open problems still remain. For instance, we do not know whether there is a monotonical (resp., anti–monotonical) behavior of co–entropy (resp., entropy) with respect to the ordering $\ll$ in all the definitions introduced in the present work.

Finally, here we focused our attention to incomplete information systems and thus to the similarity relation given in equation (5). As a future development, we plan to analyze partial partitions of the universe induced by the "random variables" generated by incomplete information systems.

References

[Ash90] R. B. Ash, *Information theory*, Dover Publications, New York, 1990, (originally published by John Wiley & Sons, New York, 1965).

[Bir67] G. Birkhoff, *Lattice theory*, third ed., American Mathematical Society Colloquium Publication, vol. XXV, American Mathematical Society, Providence, Rhode Island, 1967.

[BPA98] T. Beaubouef, F. E. Petry, and G. Arora, *Information–theoretic measures of uncertainty for rough sets and rough relational databases*, Journal of Information Sciences **109** (1998), 185–195.

[Cat98] G. Cattaneo, *Abstract approximation spaces for rough theories*, in Polkowski and Skowron [PS98], pp. 59–98.

[CC06] G. Cattaneo and D. Ciucci, *Some methodological remarks about categorical equivalence in the abstract approach to roughness. Part I*, Lecture Notes in Artificial Intelligence, vol. 4062, Springer–Verlag, 2006, pp. 277–283.

[Har28] R. V. L. Hartley, *Transmission of information*, The Bell System Technical Journal **7** (1928), 535–563.

[HHZ04] B. Huang, X. He, and XZ. Zhong, *Rough entropy based on generalized rough sets covering reduction*, Journal of Software **15** (2004), 215–220.

[Khi57] A. I. Khinchin, *Mathematical foundations of information theory*, Dover Publications, New York, 1957, (translation of two papers appeared in Russian in *Uspekhi Matematicheskikh Nauk*, **3** 1953, 3–20 and **1**, 1965, 17–75).

[KPPS99] 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.

[KW98] G. J. Klir and M. J. Wierman, *Uncertainty based information*, Physica–Verlag, New York, 1998.

[LS04] 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.

[LX00] J. Liang and Z. Xu, *Uncertainty measure of randomness of knowledge and rough sets in incomplete information systems*, Intelligent Control and Automata **4** (2000), 2526–2529, Proc. of the 3rd World Congress on Intelligent Control and Automata.

[Orl85] E. Orlowska, *A logic of indiscernibility relations*, Lecture Notes in Computer Sciences **208** (1985), 177–186.

[Paw81] Z. Pawlak, *Information systems - theoretical foundations*, Information Systems **6** (1981), 205–218.

[Paw82] ———, *Rough sets*, Int. J. Inform. Comput. Sci. **11** (1982), 341–356.

[Paw91] ———, *Rough sets: Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Dordrecht, 1991.

[PS98] L. Polkowski and A. Skowron (eds.), *Rough sets in knowledge discovery 1*, Physica–Verlag, Heidelberg, New York, 1998.

[Rez94] F. M. Reza, *An introduction to information theory*, Dover Publications, New York, 1994, (originally published by Mc Graw-Hill, New York, 1961).

[Sha48] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379–423, 623–656.

[Sle02] D. Slezak, *Approximate entropy reducts*, Fundamenta Informaticae **53** (2002), 365–390.

[SW01] D. Slezak and J. Wroblewski, *Applications of normalized decision measures to the new case classification*, Lecture Notes in Artificial Intelligence, vol. 2005, Springer-Verlag, 2001, pp. 553–560.

[Vak91] D. Vakarelov, *A modal logic for similarity relations in Pawlak knowledge representation systems*, Fundamenta Informaticae **XV** (1991), 61–79.

[Wie99] M.J. Wierman, *Measuring uncertainty in rough set theory*, International Journal of General Systems **28** (1999), 283–297.

[YLLL94] Y. Yao, X. Li, T. Lin, and Q. Liu, *Representation and classification of rough set models*, Conference Proceeding of Third International Workshop on Rough Sets and Soft Computing (San Jose, California), November 10-12 1994, pp. 630–637.

[Zak83] W. Zakowski, *Approximations in the space* (U, π), Demonstratio Mathematica **XVI** (1983), 761–769.

Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano–Bicocca, Via Bicocca degli Arcimboldi 8, I–20126 Milano (Italy)

E-mail address: {bianucci,cattang,ciucci}@disco.unimib.it