

Article

Classifying Entropy Measures

Angel Garrido

Fundamental Mathematics Department, Faculty of Sciences UNED, Paseo Senda del Rey 9.
28040-Madrid, Spain; E-Mail: agarrido@mat.uned.es

Received: 27 April 2011; in revised form: 6 July 2011 / Accepted: 6 July 2011 /

Published: 20 July 2011

Abstract: Our paper analyzes some aspects of Uncertainty Measures. We need to obtain new ways to model adequate conditions or restrictions, constructed from vague pieces of information. The classical entropy measure originates from scientific fields; more specifically, from Statistical Physics and Thermodynamics. With time it was adapted by Claude Shannon, creating the current expanding Information Theory. However, the Hungarian mathematician, Alfred Rényi, proves that different and valid entropy measures exist in accordance with the purpose and/or need of application. Accordingly, it is essential to clarify the different types of measures and their mutual relationships. For these reasons, we attempt here to obtain an adequate revision of such fuzzy entropy measures from a mathematical point of view.

Keywords: mathematical analysis; measure theory; fuzzy systems; information theory

1. Introduction

The *Shannon Entropy* is a measure of the average information content one is missing when one does not know the value of the random variable. This concept proceeds from the famous Shannon paper [1]. It represents an absolute limit on the best possible lossless compression of any communication under certain constraints, treating messages to be encoded as a sequence of independent and identically distributed random variables.

Usually, we define the Shannon Entropy by the following expression:

$$H(P) = -\sum_i p_i \log p_i = \sum_i p_i \log (1/p_i)$$

H_n will be a function of n non-negative random variables that add up to 1, and represent probabilities. H_n acts on the n -tuple of values on the sample, $(p_i)_{i=1,2,\dots,n}$.

The information that we receive from an observation is equal to the degree to which uncertainty is reduced.

Among its main *properties*, we have:

Continuity. The measure H should be *continuous*, in the sense that changing the values of the probabilities by a very small amount, should only change the H value by a small amount.

Maximality. The measure H will be maximal, if all the outcomes are equally likely, *i.e.*, the uncertainty is highest when all the possible events are equiprobable; thus,

$$H_n(p_1, p_2, \dots, p_n) \leq H_n(1/n, 1/n, \dots, 1/n)$$

And the entropy will increase with the number of outcomes,

$$H_n(1/n, 1/n, \dots, 1/n) < H_{n+1}(1/(n+1), 1/(n+1), \dots, 1/(n+1))$$

Additivity. The amount of entropy should be independent of how the process is considered, as being divided into parts. Such a functional relationship characterizes the entropy of a system with respect to the sub-systems. It demands that the entropy of every system can be identified and, then, computed from the entropies of their sub-systems.

i.e., if $S = \cup_{i=1,2,\dots,n} S_i$, then $H(S) = \sum_{i=1,2,\dots,n} H(S_i)$.

This is because *statistical entropy* is a probabilistic measure of uncertainty, or ignorance about data, whereas *Information* is a measure of a reduction in that uncertainty.

Entropy and related information measures provide descriptions of the long term behavior of random processes [2], and that this behavior is a key factor in developing the Coding Theorems of IT (Information Theory).

The contributions of Andrei Nikolaievich Kolmogorov (1903–1987) to this mathematical theory provide great advances to the Shannon formulations, proposing a new complexity theory, now translated to Computer Sciences. According to such theory, the complexity of a message is given by the size of the program necessary to enable the reception of such a message. From these ideas, Kolmogorov analyzes the entropy of literary texts and the subject Pushkin poetry. Such entropy appears as a function of the semantic capacity of the texts, depending on factors such as their extension and also the flexibility of the corresponding language.

It should also be mentioned that Norbert Wiener (1894–1964), considered the founder of Cybernetics, who in 1948 proposed a similar vision for such a problem. However, the approach used by Shannon differs from that of Wiener in the nature of the transmitted signal and in the type of decision made by the receiver.

In the Shannon model, messages are firstly encoded, and then transmitted, whereas in the Wiener model the signal is communicated directly through the channel without need of being encoded.

Another measure conceptualized by R. A. Fisher (1890–1962), the so called *Fisher Information (FI)*, applies statistics to estimation, representing the amount of information that a message carries concerning an unobservable parameter.

Certainly the initial studies on IT were undertaken by Harry Nyquist (1889–1976) in 1924, and later by Ralph Hartley (1888–1970), who in 1928 recognized the logarithmic nature of the measure of information. This was later essential the key in Shannon and Wiener’s papers.

The contribution of the Romanian mathematician and economist Nicholas Georgescu-Roegen (1906–1994), who studied in London with Karl Pearson, is also very interesting, whose great work was *The Entropy Law and the Economical Process*. In this memorable book, he proposed that the second law of thermodynamics also governs economic processes. Such ideas permitted the development of some new fields, such as Bioeconomics or Ecological Economics.

Also some others should be noted, studying a different kind of measure, the so called *inaccuracy measure*, involving two probability distributions.

R. Yager [3], and M. Higashi and G. J. Klir [4] showed the entropy measure as the difference between two fuzzy sets. More specifically, this is the difference between a fuzzy set and its complement, which is also a fuzzy set.

The *Shannon Entropy* is a measure of the average information content one is missing when one does not know the value of the random variable. These ideas proceed from their famous seminal paper [1]. It represents an absolute limit on the best possible lossless compression of any communication, under constraints, treating messages to be encoded as a sequence of independent and identically distributed random variables. The information that we receive from an observation is equal to the degree to which uncertainty is reduced. So,

$$I = H(\text{before}) - H(\text{after})$$

Finally, we may define *I(Information)* in terms of the probability, p , by the following

Properties of Information Measure, I.

- (1) $I(P) \geq 0$, *i.e.*, information is a non-negative quantity;
- (2) $I(1) = 0$, *i.e.*, if an event has probability 1, we get no information from the occurrence of the event;
- (3) If two independent events occur, the information we get from observing the events is the sum of both informations;
- (4) Information measure must be continuous, and also a monotonic function of the probability. So, slight changes in probability should result in slight changes in the information.

2. Graph Entropy

Graph theory has emerged as a primary tool for detecting numerous hidden structures in various information networks, including Internet graphs, social networks, biological networks, or more generally, any graph representing relations in massive data sets. Analyzing these structures is very useful to introduce concepts such as Graph Entropy and Graph Symmetry.

We consider a functional on a graph, $G = (V, E)$, with \mathbf{P} a probability distribution on its node set, V , and we suppose varying random samples, $P = (p_i)_{i=1,2,\dots,n}$, on the probabilistic space.

The mathematical construct called a Graph Entropy will be denoted by GE. It will be defined as

$$H(G, P) = \min_p [\sum_{i=1,2,\dots,n} p_i \log p_i]$$

Observe that such a function is convex. It tends to $+\infty$ on the boundary of the non-negative orthant of \mathbf{R}^n . Also, monotonically tends to $-\infty$ along rays departing from the origin. So, such a minimum is always achieved and will be finite.

The entropy of a system represents the amount of uncertainty one observer has about the state of the system. The simplest example of a system will be a random variable, which can be shown by a node in the graph, being their edges representative of the mutual relationship between them. Information measures the amount of correlation between two systems, and reduces in entropies to a mere difference. So, *the Entropy of a Graph* (will be denoted by GE) *is a measure of graph structure, or lack of it.*

Therefore, it may be interpreted as the amount of Information, or the degree of “surprise”, communicated by a message. As the basic unit of Information is the bit, Entropy also may be viewed as the number of bits of “randomness” in the graph, verifying that the higher the entropy, the more random the graph.

Let G now be an arbitrary finite rooted Directed Acyclic Graph (or DAG, in acronym). For each node, v , we denote $i(v)$ the number of edges that terminates at v . Then, the Entropy of the graph is definable as

$$H(G) = \sum [i(v) - 1] \log_2 [((\text{Card}(E) - \text{Card}(V) + 1)/(i(v) - 1))]$$

$H(X)$ may be interpreted in different ways. For instance, given a random variable, X , it informs us about how random X is, how uncertain we should be about X , or how much variability X has.

In a variant of the Graph Coloring Problem, we take the objective function to minimize the Entropy of such coloring. So, it is called the Minimum Entropy Coloring. In Chromatic Entropy, we understand the minimum Entropy of a coloring. Its role is essential in the problem of coding. If we consider this problem from a computational viewpoint, it will be of NP-hard type; for instance, on Interval Graphs.

The study of different concepts of Entropy will be very interesting, and not only on Physics, but also on Information Theory, and other Mathematical Sciences, considered in its more general vision. Also it may be a very useful tool for Biocomputing, for instance, or in many others, such as studying Environmental Sciences. This is because, among other interpretations with important practical consequences, the law of Entropy means that energy cannot be fully recycled.

Many quotations have been made until now referring to the content and significance of this fuzzy measure, for example:

“Gain in Entropy always means loss of Information, and nothing more” [5].

“Information is just known Entropy. Entropy is just unknown Information” [6].

Mutual Information and Relative Entropy, also called *Kullback-Leibler divergence*, among other related concepts, have been very useful in Learning Systems, both in supervised and unsupervised cases.

We attempt to analyze the mutual relationship between the distinct types of entropies, such as:

- The *Quantum Entropy*, also called Von Neumann Entropy;
- The *KS-Entropy* (from Kolmogorov and Sinai), which is also called *Metric Entropy* [7,8];
- The *Topological Entropy*; or

- The *Graph Entropy*, among others.

3. Quantum Entropy

This entropy was first defined by the Hungarian mathematician Janos Neumann (a.k.a. John von Neumann) in 1927, with the purpose of showing the irreversible behavior of quantum measurement processes. In fact, *Quantum Entropy* (will be denoted as *QE*) is an extension of the precedent Gibbs Entropy to the quantum realm [8]. It will be interpreted as the average information the experimenter obtains when he makes many copies of a series of observations, on an identically prepared mixed state. It plays a very important role in studying correlated systems, and also for defining entanglement measures. Recall that “Entanglement” is one of the properties of Quantum Mechanics that caused Einstein to dislike this theory. However, from then on, Quantum Mechanics has reached high success predicting experimental results and has also been proven on the correlation predicted by the theory of such entanglement.

We can apply the notion of QE to Networks. As QE is defined for quantum states, we need a method to map graphs into states. Such states for a quantum mechanical system are described by a density matrix. Usually, it is denoted as ρ . It is a positive semi-definite matrix with unitary trace ($\rho = 1$). There are many different ways, however, to associate graphs to density matrices. Until now, we have eliminated several problems through certain interesting results, but many open questions still remain.

Between the known results, we can see that the entropy for a d -regular graph tends to be in the limit when $n \rightarrow \infty$ to the entropy of K_n , i.e., the *n-complete n-graph*.

Another result may be that the entropy of graphs increases as a function of the cardinality of their edges.

Between the open problems, we can list some of them as relative to an interesting tool, a related matrix, called the *Normalized Laplacian*. This is defined by

$$\mathcal{L}(G) = \Delta^{-1/2} L(G) \Delta^{-1/2}$$

The *Combinatorial Laplacian Matrix* of G (abridged, *Laplacian of G*) is given as

$$L(G) = \Delta(G) - A(G)$$

Hence being computable as the difference between matrix degree, $\Delta(G)$, and adjacency matrix, $A(G)$.

The *degree* of a node, v , is the number of edges adjacent to v . Usually it is denoted by $d(v)$. The *degree sum* of the graph G is d_G , and it will be given by $d_G = \sum d(v)$. The *average degree* of G is expressed as $d_G^* = m \sum d(v)$, where m is the number of non-isolated nodes.

A graph, G , is *d-regular*, if $d(v) = d$, for all $v \in V(G)$.

The *degree matrix* of G is a $(n \times n)$ -matrix with entries given as

$$[\Delta(G)](u, v) = d(v), \text{ if } u = v; \text{ and otherwise } 0$$

So, the *Laplacian* of a graph, G , scaled by its degree-sum is a density matrix,

$$\rho_G = ((L(G))/(d_G)) = ((L(G))/(tr(\Delta(G)))) = ((L(G))/(m d_G^*))$$

With the well-known expression for the entropy of a density matrix, ρ , by $S(\rho) = -tr(\rho \log_2 \rho)$. Hence, departing from the concept of Laplacian of a Graph, we can say that $S(\rho_G)$ is the QE of G . If we suppose two decreasing sequences of eigenvalues of $L(G)$ and ρ_G , respectively given by

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = 0, \text{ and } \mu_1 \geq \mu_2 \geq \dots \geq \mu_n = 0$$

mutually related by a scaling factor, *i.e.*,

$$\mu_i = ((\lambda_i)/(d_G)) = ((\lambda_i)/(m d_G^*))$$

Therefore, the *Entropy of a density matrix* ρ_G can also be written as

$$S(G) = -\sum \mu_i \log_2 \mu_i$$

with the notational convention $0 \log 0 = 0$. Since its rows sum up to 0, we can conclude that the smallest eigenvalue of the density matrix must also be equal to zero, and the number of connected components of the graph is given by the multiplicity of 0 as an eigenvalue.

The QE is a very useful tool for problems such as when it is applied to the Enumeration of Spanning Trees.

4. Algorithmic Entropy

Algorithmic Entropy is the size of the smallest program that generates a string. It is denoted by $K(x)$, or AE. It receives many different names, for instance, Kolmogorov-Chaitin Complexity, or only Kolmogorov Complexity, also Stochastic Complexity, or Program-size complexity [9,10].

AE is a *measure of the amount of information in an object, x*. Therefore, it also measures its randomness degree. The AE of an object is a measure of the computational resources needed to specify such an object. *i.e.*, the AE of a string is the length of the shortest program that can produce this string as its output. So, the Quantum Algorithmic Entropy (QAE), also called *Quantum Kolmogorov Complexity* (QKC) is the length of the shortest quantum input to a Universal Quantum Turing Machine (UQTM) that produces the initial “qubit” string with high fidelity. Hence, the concept is very different of the Shannon Entropy, because, whereas this will be based on probability distributions, the AE is based on the size of programs.

All strings used may be elements of $\Sigma^* = \{0,1\}^*$, being ordered lexicographically. The length of a string x is denoted by $|x|$.

Let U be a fixed prefix-free Universal Turing Machine. For any string x of $\Sigma^* = \{0,1\}^*$, the *Algorithmic Entropy of x* will be defined by

$$K(x) = \min_p \{|p| : U(p) = x\}$$

From this concept, we can introduce the *t-time-Kolmogorov Complexity*, or t-time-bounded algorithmic entropy [11].

For any time constructible t , we introduce a refinement by

$$K^t(x) = \min_p \{|p| : U(p) = x, \text{ in at most } t(|x|) \text{ steps}\}$$

From these, we may obtain that for all x and y ,

$$(i) K(x) \leq K^t(x) \leq |x| + O(I), \quad (1)$$

and also

$$(ii) K^t(x/y) \leq K^t(x) + O(I)$$

The Kolmogorov-Chaitin (KC, by acronym) as new tool possesses many applications, in fields as diverse as Combinatorics, Graph Theory, Analysis of Algorithms, or Learning Theory, among others [10,11].

5. Metric Entropy

We consider now the Metric Entropy, also called *Kolmogorov Entropy*, or Kolmogorov-Sinai Entropy, in acronym *K-S Entropy*. Its name is associated with Andrei N. Kolmogorov, and his disciple, Yakov Sinai [4].

Let (X, Ω, μ) be a probability space, or in a more general way, a fuzzy measurable space [12]. Recall that a measurable partition of X is such that each one of their elements is a measurable set, therefore, an element of the fuzzy σ -algebra, Ω . And let I^X be the set of mappings from X to the closed unit interval, $I = [0,1]$.

A fuzzy σ -algebra, Σ , on a nonempty set, X , is a subfamily of I^X satisfying that

- (1) $I \in \Sigma$;
- (2) If $\alpha \in \Sigma$, then $1 - \alpha \in \Sigma$;
- (3) If $\{\alpha_i\}$ is a sequence in Σ , then $\bigvee \alpha_i = \sup_i \alpha_i \in \Sigma$;

A fuzzy probability measure, on a fuzzy σ -algebra, Σ , is a function

$$m: \Sigma \rightarrow [0,1]$$

which holds

- [1] $m(I) = 1$
- [2] for all $\alpha \in \Sigma$, $m(1 - \alpha) = 1 - m(\alpha)$
- [3] for all $\alpha, \beta \in \Sigma$, $m(\alpha \vee \beta) + m(\alpha \wedge \beta) = m(\alpha) + m(\beta)$
- [4] If $\{\alpha_i\}$ is a sequence in Σ , such that $\alpha_i \uparrow \alpha$, being $\alpha \in \Sigma$, then $m(\alpha) = \sup \{m(\alpha_i)\}$

We call (X, Ω, μ) a *fuzzy-probability measure space*, and the elements of Ω are called *measurable fuzzy sets*.

The notion of “*fuzzy partition*” was introduced by E. Ruspini. Given a finite measurable partition, \wp , we can define its *Entropy* by

$$H_\mu(\wp) = \sum_{p \in \wp} -\mu(p) \log \mu(p)$$

As usual in these cases, we take as convention that $0 \log 0 = 0$.

Let $T: X \rightarrow X$ be a measure-preserving transformation. Then, the *Entropy of T w.r.t. a finite measurable partition, \wp* , is expressed as

$$h_\mu(T, \wp) = \lim_{n \rightarrow \infty} H_\mu(V^{-k} \wp)$$

with H_μ the entropy of a partition, and where \vee denotes the join of partitions. Such a limit always exists.

Therefore, we may define the *Entropy of T* as

$$h_\mu(T) = \sup_{\mathcal{P}} h_\mu(T, \mathcal{P})$$

by taking the supremum over all finite measurable partitions.

Many times $h_\mu(T)$ is named the *Metric Entropy of T*. So, we may differentiate this mathematical object from the well-known as Topological Entropy.

We may investigate the mutual relationship of the Metric Entropy and the Covering Numbers.

Let (X, d) be a metric space, and let $Y \subseteq X$ be a subset of X . We say that $Y^* \subseteq X$ is an ε -cover of Y , if for each $y \in Y$, there exists a $y^* \in Y^*$ such that $d(y, y^*) \leq \varepsilon$. It is clear that there are many different covers of Y . But we are especially interested here in one which contains the lesser number of elements.

We call the cardinal, or size, of such a cover its *Covering Number*. Mathematically expressed, the ε -covering number of Y is

$$N(\varepsilon, Y, d) = \min \{ \text{card}(Y^*) : Y^* \text{ is an } \varepsilon\text{-cover} \}$$

A *proper cover* is one where $Y^* \subseteq Y$. And a *proper covering number* is defined in terms of the cardinality of the minimum proper cover. Both, covering numbers and proper covering numbers, are related by

$$N(\varepsilon, Y) \leq N_{\text{proper}}(\varepsilon, Y) \leq N(\varepsilon/2, Y)$$

Furthermore, we recall that the *Metric Entropy*, $H(\varepsilon, Y)$, is a natural representation of the cardinal of the set of bits needed to send, in order to identify an element of the set up to precision ε . It will be expressed by

$$H(\varepsilon, Y) = \log N(\varepsilon, Y)$$

In a dynamical system, the metric entropy is equal to zero for non-chaotic motion. And it is strictly greater than zero for chaotic motion. So, it will be interpreted as a simple indicator of the complexity of a dynamical system.

6. Topological Entropy

Let (X, d) be a compact metric space, and let $f: X \rightarrow X$ be a continuous map. For each $n > 0$, we define a new metric, d_n , by

$$d_n(x, y) = \max \{ d(f^i(x), f^i(y)) : 0 \leq i < n \}$$

Two points, x and y , are close with respect to this metric, if their first n iterates (given by $f^i, i = 1, 2, \dots$) are close.

For $\varepsilon > 0$, and $n \in \mathbb{N}^*$, we say that $S \subset X$ is an (n, ε) -separated set, if for each pair, x, y , of points of S , we have $d_n(x, y) > \varepsilon$. Denote by $N(n, \varepsilon)$ the maximum cardinality of a (n, ε) -separated set. It must be finite, because X is compact. In general, this limit may exist, but it could be infinite. A possible interpretation of this number is as a measure of the average exponential growth of the number of distinguishable orbit segments. So, we could say that the higher the topological entropy is, the more essentially different orbits we have [2,7].

From an analytical viewpoint, the topological entropy is a continuous and monotonically increasing function.

$N(n, \varepsilon)$ shows the number of “distinguishable” orbit segments of length n , assuming we cannot distinguish points that are less than ε apart.

The *topological entropy of f* is then defined by

$$H_{top} = \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} [(1/n) \log N(n, \varepsilon)]$$

Therefore, TE is a non-negative number measuring the complexity degree of the system. So, it gives the exponential growth of the cardinality for the set of distinguished orbits, according to time advances [13–16].

7. Chromatic Entropy

A system can be defined as a set of components functioning together as a whole. A systemic point of view allows us to isolate a part of the world, and so, we can focus on those aspects that interact more closely than others. The entropy of a system represents the amount of uncertainty one observer has about the state of the system [10,12]. The simplest example of a system will be a random variable, which can be shown by a node in the graph being their edges representative of the mutual relationship between them. Information measures the amount of correlation between two systems, and reduces to a mere difference between entropies. So, the Entropy of a Graph (will be denoted by GE) is a measure of graph structure, or lack of it. Therefore, it may be interpreted as the amount of Information, or the degree of “surprise”, communicated by a message. Further, as the basic unit of Information is the bit, Entropy also may be viewed as the number of bits of “randomness” in the graph, verifying that the higher the entropy, the more random the graph.

We consider a functional on a graph, $G = (V, E)$, with P a probability distribution on its node (or vertex) set, V . This mathematical construct will be denoted by GE and defined as

$$H(G, P) = \min \sum p_i \log p_i$$

Let G now be an arbitrary finite rooted Directed Acyclic Graph (DAG, in acronym). For each node, v , we denote $i(v)$ the number of their edges that terminates at v . Then, the Entropy of the graph is

$$H(G) = \sum [i(v) - 1] \log_2 [((\text{Card}(E) - \text{Card}(V) + 1)/(i(v) - 1))]$$

$H(X)$ may be interpreted in some different ways. For instance, given a random variable, X , it informs us about how random X is, how uncertain we should be about X , or how much variability X possesses.

In a variant of the “Graph Coloring Problem”, we take the objective function to minimize the Entropy of such coloring. So, it is called the *Minimum Entropy Coloring*.

In *Chromatic Entropy*, we understand the minimum Entropy of a coloring. Its role is essential in the problem of coding. If we consider this problem from a computational viewpoint, it is NP-hard; for instance, on Interval Graphs.

8. Mutual Relationship between Entropies

In the mid 1950s, the Russian mathematician Andrei N. Kolmogorov imported Shannon's probabilistic notion of entropy into the theory of dynamical systems, and showed how entropy can be used to tell whether two dynamical systems are non-conjugate, *i.e.*, non-isomorphic. His work inspired a whole new approach in which entropy appears as a numerical invariant of a class of dynamical systems. Because the Kolmogorov's metric entropy is an invariant of measure theoretical dynamical systems, it is therefore closely related to Shannon's source entropy [14].

Ornstein showed that metric entropy suffices to completely classify two-sided Bernoulli processes, a basic problem which for many decades appeared completely intractable. Recently, has been shown how to classify one-sided Bernoulli processes; this turns out to be quite a bit harder. In 1961, Adler et al. introduced [17,18] the aforementioned topological entropy, which is the analogous invariant for topological dynamical systems. There exists a very simple relationship between these quantities, because maximizing the metric entropy over a suitable class of measures defined on a dynamical system, gives its topological entropy. The relationship between TE and the Entropy in the sense of Measure Theory (K-S) is given by the so-called *Variational Principle*, which established that

$$h(T) = \sup \{h_{\mu}(T)\}_{\mu \in P(X)}$$

This may be interpreted as TE is equal to the supremum of Kolmogorov-Sinai (or K-S) entropies, $h_{\mu}(T)$, with μ belonging to the set of all T -invariant Borel probability measures on X .

The mutual relationship between Algorithmic Entropy and Shannon Entropy is that the expectation of the former gives us the latter, up to a constant depending on the distribution.

Also we may express, departing of $P(x)$ as a recursive probability distribution, that

$$0 \leq \sum P(x) K(x) - H(P) \leq K(P)$$

Finally, we recall that given a random variable, X , its *Shannon Entropy* is given by

$$H(X) = -\sum P(x) \log_2 P(x)$$

whereas the Rényi Entropy of order $\alpha \neq 1$ of such random variable will be

$$H_{\alpha}(X) = (1/(1 - \alpha)) \log_2 (\sum P(x)^{\alpha})$$

The Rényi Entropy of order α converges to the Shannon Entropy, when α tends to one, *i.e.*,

$$\lim_{\alpha \rightarrow 1} \{(1/(1 - \alpha)) \log_2 (\sum P(x)^{\alpha})\} = -\sum P(x) \log_2 P(x)$$

Hence,

$$\lim_{\alpha \rightarrow 1} H_{\alpha}(X) = H(X)$$

Therefore, the Rényi Entropy may be considered as a generalization of the Shannon Entropy, or dually, the Shannon Entropy will be a particular case of Rényi Entropy [13,14].

9. Graph Symmetry

As we know, Symmetry in a system means invariance of its elements under a group of transformations. When we take Network Structures, it means invariance of adjacency of nodes under the permutations on node set.

Let G and H be two graphs. An isomorphism from G to H will be a bijection between the node sets of both graphs, *i.e.*, a $f: G \rightarrow H$, such that any two nodes, u and v , of G are adjacent in G if and only if $f(u)$ and $f(v)$ are also adjacent in H . Usually, it is called “edge-preserving bijection”. If an isomorphism exists between two graphs, G and H , then such graphs are called *Isomomorphic Graphs*.

The graph isomorphism is an equivalence, or equality, as relation on the set of graphs. Therefore, it partitions the class of all graphs into equivalence classes. The underlying idea of isomorphism is that some objects have the same structure, if we omit the individual character of their components. A set of graphs isomorphic to each other is denominated an isomorphism class of graphs.

An automorphism of a graph, $G = (V, E)$, will be an isomorphism from G onto itself. So, a graph-automorphism of a simple graph, G , is simply a permutation on the set of its nodes, $V(G)$, $f: G \rightarrow G$, such that the image of any edge of G is always an edge in G . That is, if $e = \{u, v\} \in V(G)$, then $f(e) = \{f(u), f(v)\} \in V(G)$. Either expressed in group theoretical way, we have

$$u \sim v \text{ if and only if } ug \sim vg \text{ if and only if } u^g \sim v^g$$

The family of all automorphisms of a graph G is a permutation group on $V(G)$. The inner operation of such group is the composition of permutations. Its name is very well-known, the *Automorphism Group of G* , and abridgedly, it is denoted by $\text{Aut}(G)$. Conversely, all groups may be represented as the automorphism group of a connected graph. The automorphism group is an algebraic invariant of a graph. So, we can say that an automorphism of a graph is a form of symmetry in which the graph is mapped onto itself while preserving the edge-node connectivity. Such automorphic tool may be applied both on Directed Graphs (DGs) and on Undirected Graphs (UGs).

We will say either *graph invariant* or graph property, when it depends only of the abstract structure, not on graph representations, such as particular labeling or drawing of the graph. So, we may define a graph property as every property that is preserved under all their possible isomorphisms of the graph. Therefore, it will be a property of the graph itself, not depending on the representation of the graph.

The semantic difference also consists in its quantitative or qualitative character. For instance, when we said that “the graph does not possess directed edges”, this will be a property, because it is a qualitative statement. While when we say “the number of nodes of degree two in such graph”, this would be an invariant, because it is a quantitative statement.

From strictly a mathematical viewpoint, a graph property can be interpreted as a class of graphs, composed by graphs that have the accomplishment of some conditions in common. Hence, a graph property can also be defined as a function of whose domain would be the set of graphs, and its range will be the bi-valued set composed of two options, true and false, $\{T, F\}$, according to which a determinate condition is either verified or violated for the graph. A graph property is called *hereditary*, if it is inherited by its induced subgraphs. And it is additive, if it is closed under disjoint union. For

example, the property of a graph to be planar is both additive and hereditary. Instead of this, the property of being connected is neither.

The computation of certain graph invariants may be very useful for the purpose of discriminating when two graphs are isomorphic, or rather non-isomorphic. The support of these criteria will be that for any invariant at all, two graphs with different values cannot be isomorphic between them. However, two graphs with the same invariants may or may not be isomorphic between them. So, we will arrive to the notion of completeness.

Let $I(G)$ and $I(H)$ be invariants of two graphs, G and H . It will be considered complete if the identity of the invariants ever implies the isomorphism of the corresponding graphs, *i.e.*, if $I(G) = I(H)$, then G will be isomorphic to H .

A directed graph, or *digraph*, is the usual pair $G = (V, E)$, but now with an additional condition: it has at most one directed edge from node i to node j , being $1 \leq i, j \leq n$. We add the term “*acyclic*” when there are no cycles of any length. Usually, we use the acronym DAG to denote an acyclic directed graph. A very important result may be this: For each n , the cardinality of the n -DAGs, or DAGs with n labeled nodes, is equal to the number of $(n \times n)$ -matrices of 0's and 1's whose eigenvalues are positive real numbers.

It is possible to prove that every group is the automorphism group of a graph. If the group is finite, the graph may be taken to be finite. Further, George Pólya observed that not every group must be the automorphism group of a tree.

10. Symmetry as Invariance

One of the more fundamental results in Physics and in any Science [12,14—16,19] is that obtained by the great mathematician *Emmy Noether* (1882–1935). This was proved in 1915, and published in 1918. It states that any differentiable symmetry of the action of a physical system has a corresponding conservation law. Hence, for each continuous symmetry of a physical theory there is a corresponding conserved quantity, *i.e.*, a physical quantity that does not change with time. So, Symmetry under translation corresponds to conservation of momentum; Symmetry under rotation to conservation of angular momentum; Symmetry in time to conservation of energy. Also it is present in Relativity Theory, Quantum Mechanics and so on. It is a very important result, because it allows us to derive conserved quantities from the mathematical form of our theories. Recall that the action of a physical system is an integral of a Lagrangian function, from which the behavior of the system can be determined by the Principle of Least Action. Note that this theorem does not apply to systems that cannot be modeled with a Lagrangian, for instance to dissipative systems.

The *Noether Theorem* has become essential not only in modern Theoretical Physics, but in the Calculus of Variations, and therefore, in fields such as Modeling and Optimization. In fact, all modern Physics is based on a bunch of Symmetry Principles, from which the rest follows. So, we can say that *the Laws of Nature are constrained by Symmetry*. Such theorem admits distinct but essentially equivalent, statements, as may be “*to every differentiable symmetry generated by local actions, there corresponds a conserved current*”. This connects today with many evolving subjects of modern Physics, such as Gauge Symmetry, in Quantum Mechanics, the results of Witten (String Theory) and many others. Noether is remembered not only by this theorem (actually, they are two results, with

many consequences), but by many contributions to Abstract Algebra. There is also a quantum version of this Noether's theorem, known as the Ward-Takahasi Identity.

The conservation law of a physical quantity is expressed by a continuity equation, where the conserved quantity is named Noether's Charge, and the flow carrying that "charge" is the Noether's Current. In Quantum Mechanics, invariance under a change of phase of the wave function leads to the Conservation of Particle Number.

11. Fuzzy Entropies

In recent decades, the expansion of fuzzy mathematics and its applications are very formidable [17,20]. The parallel version of different mathematical fields, but adapted to degrees of truth, is in advance. The basic idea according to which an element not necessarily belongs totally, or does not belong in absolute, to a set, but it can belong more or less, that is, in some degree it signifies a modern revolution in scientific thinking, adapting the sometimes hieratic mathematics to the features of the real world. So, it produces new fields, such as Fuzzy Measure Theory, which generalizes the classical Measure Theory of Lebesgue and other authors. It must be very useful as a tool in our own papers and occurs in every mathematical field. In Fuzzy Modeling we attempt to construct Fuzzy Systems. Many times, it will be a very difficult task, because it is necessary to identify many parameters. It offers a great potential for analyzing structures with non-stochastic imprecise input information.

In Fuzzy Optimization [17,21], our objective is to maximize or minimize a fuzzy set submitted to some fuzzy constraints, but we cannot make this directly with the "value" of a fuzzy set. For this reason, in areas such as Finance, we wish to maximize/minimize the value of a discrete/continue random variable, being restricted by a probability mass/density function. So, we change the multi-objective problem into a single crisp objective subject to the fuzzy constraints and it is possible to generate good approximate solutions by Genetic Algorithms. Also there are different fuzzy optimization problems, which include learning a Fuzzy Neural Network, useful to solve fuzzy linear programming problems (FLP), and fuzzy inventory control, using such Genetic Algorithms.

12. About Negentropy

Negentropy is essential for the axiomatized concept of entropy (denoted by H). Many of its seminal ideas were derived from Claude E. Shannon [1], and Alfred Rényi [17,22]. It is also related to the coding length of the random variable. In fact, with some simple assumptions, H is the coding length of the random variable. Entropy is the basic concept of Information Theory. It can be interpreted, for a random variable, as the degree of information that the observation of the variable produces. The more "randomness" presented in the variable, the larger the entropy. It is defined, for a discrete random variable, Y , as

$$H(Y) = -\sum P(Y = y_i) \log P(Y = y_i)$$

where the y_i are the possible values of Y .

This may be generalized for the continuous case, being then called *Differential Entropy* (also named *continuous entropy*). It will be defined by

$$H(y) = -\int f(y) \log f(y) dy$$

with $f(y)$ density function, associated with the continuous random variable Y .

There exists a very important result of Information Theory, according to a *Gaussian random variable has the largest entropy, among all random variables of the same variance*. So, the Normal, or Gaussian distribution is the “least structured”, or equivalently, the “most random” among all distributions. But we have a second and very important measure of non-gaussianity (departure from the Normal). It is called with distinct names, such as *Negentropy*, either Negative Entropy or *Syntropy*, denoted by J . Actually, it is a slightly modified version of differential entropy, defined by

$$J(y) = H(y_{gauss}) - H(y)$$

being y_{gauss} a Gaussian random variable of the same covariance matrix as y .

Some of its properties are interesting, as

$$J(y) \geq 0, \text{ for each } y$$

That is, Negentropy is always non-negative. And it is null in the case of the Normal distribution:

$$J = 0 \text{ if and only if it is Gaussian}$$

According to Schrödinger’s classical book *What is Life?*

“Negentropy of a living system is the entropy that it exports, to maintain its own entropy low”

And Brillouin [23] says that

“A living system imports negentropy, and stores it”

The Curie Principle of Symmetry, due to Pierre Curie, postulates that *the symmetry group of the cause is a subgroup of the symmetry group of the effect*. This idea may produce deep ramifications on Causality Theory, and also analyzing relationships among the foundations of physical theories.

13. Conclusions

Statistical entropy is a probabilistic measure of uncertainty, or ignorance about data. However, Information should be the measure of the reduction in that uncertainty. The Entropy of a probability distribution is just the expected value of the information of such a distribution. All these improved tools must allow us to advance not only in fields such as Optimization Theory, but also on Generalized Fuzzy Measures, Economics, modeling in Biology, and so on [17,24,25]. Here, we have shown some different entropy measures, more or less useful depending on its context and their need of applications, according to ideas suggested by the Hungarian mathematician Alfred Rényi many years ago [22, 26–28].

Acknowledgements

I wish to express my gratefulness to Joe Rosen, Shu-Kun Lin, and Joel Ratsaby, which have proposed to collaborate newly with this paper in *SYMMETRY*. And also to my anonymous referees for their very wise commentaries.

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
2. Wehrl, A. General properties of entropy. *Rev. Mod. Phys.* **1978**, *50*, 221–260.
3. Yager, R. On the Measure of Fuzziness and Negation. *Int. J. General Syst.* **1979**, *5*, 221–229.
4. Higashi, M., and Klir, G.J. Measures of Uncertainty and Information based on possibility distributions. *Int. J. General Syst.* **1982**, *9*, 43–58.
5. Lewis, G. N. The Entropy of Radiation. *Proc. Natl. Acad. Sci. USA* **1927**, *13*, 307–314.
6. Frank, M. P. Approaching Physical Limits of Computing. *Multiple-Valued Logic* **2005**, doi: 10.1109/ISMVL.2005.9.
7. Simonyi, G. Graph Entropy: A Survey. *DIMACS* **1995**, *20*, 399–441.
8. Sinai, G. On the concept of Entropy of a Dynamical System. *Dokl. Akad. Nauk SSSR* **1959**, *124*, 768–771.
9. Passarini, F.; Severini, S. *The von Neumann Entropy of Networks*. University of Munich: Munich, Germany, 2009.
10. Volkenstein, M.V. *Entropy and Information (Progress in Mathematical Physics)*; Birkhäuser Verlag: Berlin, Germany, 2009; Volume 57.
11. Devine, S. The insights of algorithmic entropy. *Entropy* **2009**, *11*, 85–110.
12. Dehmer, M. Information processing in Complex Networks: Graph entropy and Information functionals. *Appl. Math. Comput.* **2008**, *201*, 82–94.
13. Jozsa, R. Quantum Information and Its Properties. In *Introduction to Quantum Computation and Information*; Lo, H.K., Popescu, S., Spiller, T., Eds.; World Scientific: Singapore, 1998.
14. Titchener, M.R.; Nicolescu, R.; Staiger, L.; Gulliver, A.; Speidel, U. Deterministic Complexity and Entropy. *J. Fundam. Inf.* **2004**, Volume 64.
15. Titchener, M.R. A Measure of Information. In *Proceedings of Data Compression Conference 2000*, Snowbird, UT, USA, 2000; pp. 353–362.
16. Titchener, M.R. A Deterministic Theory of Complexity, Information and Entropy. In *Proceedings of IEEE Information Theory Workshop*, San Diego, CA, USA, February 1998.
17. Preda, V. Balcau, C. *Entropy Optimization with Applications*. Editura Academiei Romana: București, România, 2010.
18. Li, M.; Vitányi, P. *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed.; Springer Verlag: Berlin, Germany, 2008.
19. Dumitrescu, D. Entropy of a fuzzy process. *Fuzzy Sets Syst.* **1993**, *55*, 169–177.
20. Wang, Z.; Klir, G.J. *Generalized Measure Theory*; Springer Verlag: Berlin, Germany and New York, NY, USA, 2008.
21. Garrido, A.; Postolica, V. *Modern Optimization. Editura Matrix-Rom*: Bucuresti, Romania, 2011.
22. Rényi, A. On measures of information and entropy. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, Berkeley, CA, USA, 20 June–30 July 1960; University of California Press: Berkeley, CA, USA, pp. 547–561.
23. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1956**, *108*, 171–190.

24. Georgescu-Roegen, N. *The Entropy Law and the Economic Process*; Harvard University Press. Cambridge, MA, USA, 1971.
25. Liu, X. Entropy, distance and similarity measures of fuzzy sets and their relations. *Fuzzy Sets Syst.* **1992**, *52*, 305–318.
26. De Luca, A.; Termini, S. A definition of non-probabilistic entropy, in the setting of fuzzy theory. *Inf. Control* **1972**, *20*, 301–312.
27. You, C.; Gao, X. Maximum entropy membership functions for discrete fuzzy variables. *Inf. Sci.* **2009**, *179*, 2353–2361.
28. Dumitrescu, D. Fuzzy measures and the entropy of fuzzy partitions. *J. Math. Anal. Appl.* **1993**, *176*, 359–373.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).