Abstract

We present a basic axiomatization for the discrete case of the concept of entropy by using a minimum amount of axioms. This axiomatization is satisfied by most of the entropies that are known in the literature, as for instance Shannon's, Renyi's, Varma's, etc. We analyze also the mathematical consequences of these axioms from a topological and functional point of view. This analysis leads us to the conclusion that entropies live in a non compact convex subset of the Banach space $C(\Delta_n)$ with the sup norm. Finally, we prove that there exist as many different families of entropies as real numbers. A result with immediate relevance in different fields of knowledge, ranging from statistics to engineering.

1 Introduction

In 1948 C. Shannon [1] introduced the concept of information theoretic entropy, defined by the formula

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln(p_i). \tag{1}$$

The importance of this entropy goes beyond the original scope that C.Shannon intended when he introduced it. In 1957 the physicist E. Jaynes introduced a new technique to find prior probabilities in Bayesian Statstics [2–4]. The idea was to maximize Shannon's entropy subject to different constraints. According to Jaynes the maximization of entropy principle (MEP) ought to be the best way to find prior probabilities whatever testable information we have at hand. This, despite the fact that Shannon's entropy is not the unique way to measure information content. In Jaynes' words [5, 6]

> "One...Important reason for preferring the Shannon measure of information is the only one that satisfies...[Shannon's postulates]. Therefore one expects that any deduction made from other information measures, if carried far enough, will eventually lead to a contradiction".

Time has shown that no such contraditcion arrives, in fact, uses of other information measure proves to be very valuable when studying different types of systems. For example in statistical mechanics we have the concept of Tsally's entropy [7–10]. Also a bunch of all other entropies have found their applications in areas very far from physics, like in Image analysis [11].

The previous discussion gives an idea of the relevance that a solid axiomatization of the concept of entropy would have in several branches of knowledge. This relevance is the reason why several attempts have been made to give an axiomatic foundation to the theory of entropies. These different axiomatic approaches can be found elsewhere [12–15] .

The way the axioms are chosen in order the develop the theory is a matter of taste, as long as most of the known entropy-like functions fulfill this axioms, and these axioms reflect what we intuitively understand as information.
The aim of this paper is to give a cogent axiomatic foundation for the theory of discrete entropies, trying to put as little axioms as possible. After this, we explore the consequences that the set of chosen axioms have, from the mathematical point of view and from the applications point of view.

The difference between this paper and other papers that axiomatize entropies as well (e.g. [13, 16]) is that in this work, the number of axioms is kept to a minimum, therefore obtaining more general results. Results that are applicable to a wider range and are more fundamental.

This papers goes as follows: Section 2 Presents a brief discussion on what do we mean by an information measure; Section 3 presents a brief summary of some well known generalized entropies; In section 4 we develop the consequences of the axiomatization, where the main results are presented; Finally section 5 shows the analysis and conclusions of the results obtained.

2 What does an information measure mean?

To illustrate what we mean by an information measure, consider the following example: Someone throws a dice, of course, if we knew the initial conditions when the die was thrown, we can use Newton's second law to predict the outcome of the dice. However knowing all the information necessary to predict the outcome is practically impossible, therefore we cannot say for sure which of the numbers from 1 to 6 is going to be shown by the dice. An smarter approach could be the following: Assign a probability $\{p_1, p_2, \ldots, p_6\}$ to each of the possible outcomes, using the 'information' we have about the system. The key question is. How can we assign the probabilities to the different outcomes, based on what we know? We can argue that there must be a way to assign probabilities based on what we know about the system. For example if we know nothing about the system and the initial conditions of the dice prior to be thrown, then, our intuition tells us that the logical thing to do is to assign equal probabilities to each of the outcomes i.e. $p_1 = \frac{1}{6}, \ldots, p_6 = \frac{1}{6}$. A more complicated scenario is when you know that the die is charged, say towards 4. Clearly in this case equal a prior probabilities doesn't seem to be the right way to go, but some probability distribution centered at 4.

In the previous situation one may ask. Assuming that there is a way to assign probabilities that reflects our state of knowledge about the system, is there any systematic way to assign those probabilities? Here is where the concept of information measure comes into play. In our previous example assume that there exists a bounded function f of the probabilities assigned, such that for each probability distribution $\{p_1, \ldots, p_6\}$ assigns a non negative real number , i.e. $f(p_1, \ldots, p_6) \in \mathbf{R}^+$. If that number were a quantitative representation of how much we know about the system, then, we can arbitrarily assign the number 0 to total knowledge (i.e we know the outcome before throwing the dice) and the number $\sup_x f(x)$ to total ignorance (i.e. we don't know anything about the system). A function with the characteristics described above, can be considered as an information measure.

Our goal in this paper is to find the biggest set of functions that can be considered as information measures.To do that we first propose the set of axioms that we consider that must be fulfilled by an information measure. After that, we are going to explain the intuition behind those axioms, and the practical applications that those axioms have.

Axioms

We are going to start this discussion from a purely mathematical point of view and then we are going to put things down to earth.

For any $n \in \mathbb{N}$, define the **continuous** function

$$\tilde{S} : \Delta_n \subset \mathbb{R}^n \to \mathbb{R}, \tag{2}$$

where

$$\Delta_n = \{(p_1, \ldots, p_n) \in \mathbb{R}^n | \textstyle\sum_{i=1}^n p_i = 1, p_i \geq 0\}.$$

Clearly the points in the domain of the function $\tilde{S}$ can be regarded as probabilities of a set of exhaustive events. Also it is worth noticing that this domain is nothing more than the n-simplex from algebraic topology.

We want the function $\tilde{S}$ to satisfy the following axioms

1. Given $\sigma \in P_n$ where P_n is the permutation group of a set with n elements. Let $x = (p_1, \ldots, p_n) \in \Delta_n$ then $\tilde{S}$ satisfies permutation of its arguments invariance, that is, $\tilde{S}(x) = \tilde{S}(\sigma(x))$ where $\sigma(x) = \sigma(p_1, \ldots, p_n) = (p_{\sigma(1)}, \ldots, p_{\sigma(n)})$.
2. Given x $\in \Delta_{n-1}$, then $\tilde{S}(x) = \tilde{S}(x, 0)$.
3. Define $u_k = (\frac{1}{k}, \ldots, \frac{1}{k}) \in \Delta_k$, then if $m, n \in \mathbb{N}$ with $m \leq n$, we have $\tilde{S}(u_m) \leq \tilde{S}(u_n)$. Also $\tilde{S}(x) \leq \tilde{S}(u_n)$ for all $x \in \Delta_n$.
4. $\tilde{S}(x) = 0$ if and only if x is one of the canonical base vectors of $\mathbb{R}^n$.

This four axiom are intended to reflect the minimum properties that are expected for an information measure. Lets do a little digression on the idea that every axiom is trying to convey.

Axiom 1,says that the only thing that matters in the entropy function is the probabilities per se, not which probability gets each of the event whose probability is being measured.This is reasonable, because for example suppose that we like to measure our uncertainty about a coin toss and is known that the probability of head is 0.3 and tail 0.7, we have the same amount of information if we know that the probability of head is 0.7 and of tail 0.3.

Axiom 2 tells that if in a set of events is included one event with probability 0 our state of knowledge remains the same.

Axiom 3 states the fact that equal probabilities for exhaustive events is the most ignorant we can be about something and the more events, the more ignorant we are.

Finally axiom 4 just says that if we have an event with probability 1 (implying that the probabilities of the other events are 0). We know before hand the outcome so we are not ignorant at all.

It is worth noticing that we asked the function $\tilde{S}$ to be continuous. The reason for this is the following. If the probabilities of an event change little then our state of knowledge represented by $\tilde{S}$, also changes little. This is the intuitive meaning of continuity.

Now that we stated what we want, we need to check how useful it is. To do so, we are going to review some of the well known discrete entropies that exists in the literature.

3 A brief Review on generalized entropies

To put ideas into context it is necessary to review a few of well known generalized entropies and check if they satisfy the axioms proposed in the previous section. It is an straightforward exercise to check that the axioms are fulfilled by each of the functions that we are going to present, therefore we only going to present the functional form of this entropies.

Tsallis Entropy

One of the most famous attempts to generalize Shannon's entropy was made by the Brazilian Physicist Constantino Tsallis, with the so-called Tsallis entropy [7].

$$S(p_1,\ldots,p_n) = \frac{1-\sum_{i=1}^{n} p_i^q}{q-1} \tag{3}$$

It is a generalization of Shannon's entropy in the sense that when the parameter q goes to 1, Tsallis entropy becomes Shannon's entropy.

Renyi's Entropy

Another famous generalization is Renyi's entropy [17], which has wide application in many areas of mathematics.

$$H_\alpha(p_1,\ldots,p_n) = \frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{n} p_k^\alpha\right) \tag{4}$$

Varma and Kapur

There are also some generalizations of Renyi's entropy of order α [18, 19].

$$\frac{\ln(\sum_{i=1}^{n} p_i^{r-m+1})}{m-r}, m-1<r<m, m\geq 1, \tag{5}$$

$$\frac{\ln(\sum_{i=1}^{n} p_i^{\frac{r}{m}})}{m(m-r)}, 0<r<m, m\geq 1. \tag{6}$$

$$\frac{ln(\frac{\sum_{i=1}^{n} p_i^{t+s-1}}{\sum_{i=1}^{n} p_i^{s}})}{1-t}, t\neq 1, t>0, s\geq 1. \tag{7}$$

Non-additive Measures

In statistical mechanics the concept of entropy satisfies additivity, however some systems does not satistfy this additivity condition, hence are called non-additive systems [9]. Therefore entropies that are non-additive are of great importance, like the following ones [20, 21].

$$-2^{r-1}\sum_{i=1}^{n} p_i^r \ln(p_i), r > 0, \tag{8}$$

$$\frac{\sum_{i=1}^{n} p_i^r - p_i^s}{2^{1-r} - 2^{1-s}}, r \neq s, r > 0, s > 0, \tag{9}$$

$$-2^{1-r}\frac{\sum_{i=1}^{n} p_i^r sin(\ln(p_i))}{sin(s)}. \tag{10}$$

$$-\frac{\sum_{i=1}^{n} v_i \ln(p_i)}{\sum_{i=1}^{n} v_i}, \tag{11}$$

$$\frac{ln(\frac{\sum_{i=1}^{n} p_i^{r-1} v_i}{\sum_{i=1}^{n} v_i})}{1-t}, s \neq 1, s > 0, \tag{12}$$

$$\frac{(\exp((s-1)\sum_{i=1}^{n} v_i \ln(p_i))/\sum_{i=1}^{n} v_i}{2^{s-1}-1}, s \neq 1, s > 0, \tag{13}$$

$$\frac{((\sum_{i=1}^{n} p_i^{r-1} v_i)/\sum_{i=1}^{n} v_i)^{\frac{s-1}{r-1}} - 1}{2^{s-1}-1}, r \neq 1, s \neq 1, r > 0, s > 0. \tag{14}$$

These are only a few of the generalized entropies that exists in today's literature where as far as we know there exists more than 20 of those.

Now that we have checked that our axiomatization is useful in the entropy world, it is time to explore the consequences of those axioms. This is going to be done next.

4 Mathematical consecuences of the axiomatization

Before the main results are presented,it is necessary to do some mathematical preliminaries, in order to put ideas in a more clear language.

Definition 4.1. *Let $x, y \in \Delta_n$, for any $n \in \mathbb{N}$, we say that x is equivalent to y $(x \sim y)$ if and only if $x = \sigma(y)$ for some $\sigma \in P_n$.*

It is straightforward to check that $\sim$ is an equivalence relation over Δ_n. Now we are going to state a lemma an a definition that will serve as a basis for a new construction over the information measure $\tilde{S}$.

Lemma 4.2. *Let $x = (p_1, \ldots, p_n) \in \Delta_n$ such that $p_{i_1} = p_{i_2} = \ldots = p_{i_k} = 0$ for $k < n$ then there exists $\sigma \in P_n$ such that the last k entries of $\sigma(x)$ are 0.*

Definition 4.3. *Given the equivalence relation $\sim$ on Δ_n. We call D_n the quotient set $\Delta_n / \sim$.*

The last lemma and definition shows that without lost of generality, if a vector $x \in \Delta_n$ has k 0's in it. Is safe to assume that they are in the last k entries (this vector is the class representative of the equivalence class of x). Now we are ready to define a more concrete mathematical element that can be regarded as a true information measure.

Definition 4.4. *We define the function $S : D_n \to \mathbb{R}$, by the relation $S([x]) = \tilde{S}(x)$, where $[x]$ is the equivalence class of x by the equivalence relation $\sim$.*

There is a very powerful reason to define the new function S, besides the fact that it is easy to work with class equivalences. In the four axioms, nothing was said about how the entropy must behave when measuring uncertainity on conjunction of independent events. For example if there are 3 independent events A_1, A_2, A_3 with probabilities p_1, p_2, p_3. then one of the possible vector containing the values of the probabilities of the conjuntion of all the events is (p_1p_2, p_1p_3, p_2p_3). But the vector (p_2p_3, p_1p_3, p_1p_2) has the same information about all the conjunction of the 3 events. Thanks to the definition of S there is no need to worry about which vector to use. So, this facilitates future studies on axiomatization of entropies and how they behave under composition of diferent events.

With the new domain of definition D_n, it is necessary to see which properties of the set Δ_n are inherited by the quotient set $D_n = \Delta_n / \sim$. To study this, first a topology must be define on the quotient set. The natural way to do this is by giving Δ_n / R the finest topology that makes the canonical projection function continuous [22,23]. Here is used the universal convention to denote the canonical projection by the letter π . The topology generated by π is called the quotient topology. From the topological properties of continuous function, is concluded that because Δ_n is a closed, bounded set on $\mathbb{R}^n$, then it is compact. Obviously is connected also. Then, since $D_n = \Delta_n / \sim$

is the image of π, is also compact[1] and connected. In fact it can be shown that Δ_n is homeomorphic to $\Delta_n/\sim$.

Theorem 4.5. *The sets Δ_n and $\Delta_n/\sim$ are homeomorphic.*

Proof. For the sake of simplicity consider the case of the 2-simplex which is just a triangle, as shown in figure 1. The barycenter is nothing more than the point where all entropies attains their maximum. Also observe that this is a common point for the triangles COE,EOB,BOD,DOA, AOF,and FOC . Obviously any probability vector (p_1, p_2, p_3) belongs to any of this triangles, but the following is also true: If the vector belongs to COB, for example, then his 6 permutations belongs each one to a diferent triangle. Hence all of the triangles contains 1 representative of each of the equivalent classes. This means that all of the quotient space can be thought as any of the 6 triangles. Because any of this triangles has the same dimension of the 2-simplex we can use the result that given two simplexes they are homeomorphic if they have the same dimension. Hence the proof is completed. □

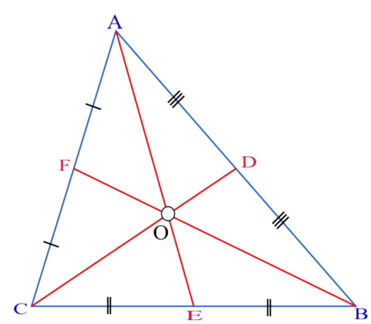

Figure 1: 2-simplex and its Barycenter

This theorem shows that working with $\tilde{S}$ or S is basically the same thing. But the advantage of working with S instead of $\tilde{S}$ is that we are working with class equivalences instead of single points in the $n-simplex$. From now on, no distinction is going to be made between a vector x and its equivalence class $[x]$. To finish the construction of S we are going to stated the following key result.

Corollary 4.6. *S is continous.*

[1]Therefore by the Wierstrass theorem, S, which is continuous as shown in theorem 4.6 attain a maximum in D. That maximum is reached in the equivalence class of the vector u_n. To see this,let [x] $\in$ D,then the following holds $S([x]) = \tilde{S}(x) \leq \tilde{S}(u_n) = S([u_n])$

Proof. Since $S = \tilde{S} \circ \pi$, and each of these functions is continuous, the result holds. $\square$

This last corollary along with the result that D_n for any $n \in \mathbb{N}$ is compact guarantee that S also attains its maximum [24]. So there is no problem about using maximization principles on the function S.

Now that all the mathematical tools have been developed and the language is clear, we can group all the entropies into one set, this set is defined as follows

Definition 4.7. *For any $n \in \mathbb{N}$ we define $X = \{S : D_n \to \mathbf{R}|\ S$ is continuous and satisfies axioms 1 to 4$\}$.*

Since entropies are continuous, X is a subset of the set of real valued continuous functions on $D_n = \Delta_n/\sim$, denoted by $C(D_n)$. Because the entropy reach its maximum on D_n, X can be regarded as a subset of the Banach space $C(D_n)$ with the sup norm. That is: given $f \in C(D_n)$, $\|f\| = \max_{x \in D_n} |f(x)|$ [25].

Whenever we have the objects under study in one particular set, we want to know more about that set, to see if we can get any insights about its structure and hopefully those insights will lead us to some real world applications.

Theorem 4.8. *The set X is not closed.*

Proof. Take the sequence $\{S_{\frac{1}{n}}\}_{n \in \mathbf{N}}$ where $S_{\frac{1}{n}}$ is the Tsallis entropy

$$S_{\frac{1}{n}} = \frac{1 - \sum_{i=1}^{m} p_i^{\frac{1}{n}}}{\frac{1}{n} - 1}. \tag{15}$$

It is straightforward to see that $\lim_{n \to \infty} S_{\frac{1}{n}} = m - 1$. However, the constant function $m - 1$ does not satisfies axiom 4. Therefore, exists sequence in X that do not converge in the set, so X does not contain all of its accumulation points, hence cannot be closed [22, 23]. $\square$

Since every compact set is closed and bounded, we conclude that the following corollary holds.

Corollary 4.9. *The set X is not compact.*

This result makes a little harder future analysis on this set, because compactness is a desirable property. The reason is that an infinite compact set behaves very much like a finite one, making analysis more easy.

Despite the fact that X is not compact, X has a very nice topological property.

Theorem 4.10. *The set X is convex.*

Proof. Let $S_1, S_2 \in X$ and let $S(x,t) = (1-t)S_1(x) + tS_2(x)$, with $x = (p_1, \ldots, p_n)$. Clearly S is invariant under permutations. Also we have that $S(x,0,t) = (1-t)S_1(x,0)+tS_2(x,0)$ and because S_1 and S_2 satisfies axiom 2, we have that $S(x,0,t) = (1-t)S_1(x,0)+tS_2(x,0) = (1-t)S_1(x)+tS_2(x) = S(x,t)$ hence $S(x,t)$ also satisfies axiom 2 for all t. Now let u_k be defined as in axiom 3, then if m$\leq$n we have that $S_i(u_m) \leq S_i(u_n)$ for i=1,2, we have that for $0 \leq t \leq 1$

$$S(u_m,t) = (1-t)S_1(u_m) + tS_2(u_m) \leq (1-t)S_1(u_n) + tS_2(u_n) = S(u_n,t). \tag{16}$$

Also is clear that the maximum of $S(x,t)$ is reached when S_1 and S_2 are evaluated at the vector u_n, therefore given $x \in \Delta_n/R$ we have $S(x,t) \leq S(u_n,t)$.Finally $(1-t), t, S_1, S_2$ is positive, so we have that if $S(x,t) = 0$ the only posibility is that $S_1 = 0$ and $S_2 = 0$ and this is true if and only if x is one of the basis vectors of $\mathbb{R}^n$. So we can conclude that $S(x,t) \in X$ for all $0 \leq t \leq 1$. Hence X is convex. □

The fact that X is a convex set allow the following important result.

Theorem 4.11. *There exists at least as much entropies as real numbers.*

Proof. The function $\varphi : [0,1] \to X$ defined by the formula

$$\varphi(t) = (1-t)S_1(x) + tS_2(x), \tag{17}$$

for fixed entropies S_1 and S_2 is a one to one mapping. Since the cardinal of real numbers is the same as the cardinal of $[0,1]$, we conclude that the cardinal of X is bigger than the cardinal of $[0,1]$. □

Finally since the set of entropies is a subset of the set of continuous functions, a well known result states that the cardinal of the set of real valued functions is the same as the cardinal of the real numbers [26]. Using the previous theorem with this fact we obtain the major mathematical result of this paper.

Theorem 4.12. *There exists as much entropies as real numbers.*

The question that arises now is. What are these results good for? A first approach to this question will be given in the conclusions section.

5 Conclusions

Because only a few axioms were used, the results presented here are very general and with wide applicability. Also shows that only 4 axioms are needed in order to convey a nice structure of the set of entropies, both from a topological point of view and a physical point of view.

To show an idea of how the results presented in this work could be useful outside the realm of mathematics, consider the first consequence of theorem 4.10. if a set of positive number $\{t_k\}_{k=1}^{n}$ are such that $\sum_{k=1}^{n} t_n = 1$, then for any set of arbitrary entropies $S_k \in X$, $k = 1, 2 \dots n$ we can create a new entropy

$$S = \sum_{k=1}^{n} t_k S_k. \tag{18}$$

The above entropy 'joins' together as many entropies as we want, and the coefficients $\{t_k\}$ are the weight on each entropy. Putting this in simple terms, we have created a way to ensemble an entropy that shares features of each of the S_k entropies and each feature is as important as the coefficient t_k. This have an inmediate application in image thresholding [11,27] where different types of entropies are used [28,29], but there exist no consensus of which entropy is better to maximize in order to achieve the best thresholding value. With the mixed entropies proposed in this text it is easy to experiment with different types of entropies and find out which one is better.

This argument can also be used in physical systems, so for example we can create a new entropy, based on old entropies, that captures the non reversibility of the system under study.

One final possible application is in the field of Bayesian statistics, where the maximum entropy principle is used to choose a objective prior probability distribution. Improving the foundations and understanding on the theory of entropies could enhanced how to find objective a prior probabilities

The possible applications described above give an idea of how the results obtained in this work can be used to produce results outside the theoretical framework presented here. Although a lot of theoretical work is needed in order to fully explore the consequences that this work brings into the field of entropies.

References

[1] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 1948.

[2] Edwin Thompson Jaynes. *ET Jaynes: Papers on probability, statistics, and statistical physics*, volume 50. Springer, 1989.

[3] Phil Gregory. *Bayesian logical data analysis for the physical sciences*, volume 10. Cambridge University Press Cambridge, 2005.

[4] Devinderjit Sivia and John Skilling. *Data analysis: a Bayesian tutorial*. Oxford University Press, USA, 2006.

[5] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[6] vaslav Brukner and Anton Zeilinger. Conceptual inadequacy of the shannon information in quantum measurements. *Physical Review A*, 63(2):022113, 2001.

[7] Constantino Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1-2):479–487, 1988.

[8] Evaldo MF Curado and Constantino Tsallis. Generalized statistical mechanics: connection with thermodynamics. *Journal of Physics A: Mathematical and General*, 24(2):L69, 1991.

[9] Constantino Tsallis. *Introduction to nonextensive statistical mechanics: approaching a complex world*. Springer, 2009.

[10] Murray Gell-Mann and Constantino Tsallis. Nonextensive entropy-interdisciplinary applications. *Nonextensive Entropy-Interdisciplinary Applications, by Edited by Murray Gell-Mann and C Tsallis, pp. 440. Oxford University Press, Apr 2004. ISBN-10: 0195159764. ISBN-13: 9780195159769*, 1, 2004.

[11] Jagat Narain Kapur and Hiremaglur K Kesavan. *Entropy optimization principles with applications*. Academic Press, 1992.

[12] W Ochs. A new axiomatic characterization of the von neumann entropy. *Reports on Mathematical Physics*, 8(1):109–120, 1975.

[13] Imre Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10(3):261–273, 2008.

[14] Bernhard Baumgartner. Characterizing entropy in statistical physics and in quantum information theory. *arXiv preprint arXiv:1206.5727*, 2012.

[15] János Aczél and Zoltán Daróczy. On measures of information and their characterizations. *New York*, 1975.

[16] Frederic Dupuis, L Kraemer, Philippe Faist, Joseph M Renes, and Renato Renner. Generalized entropies. *arXiv preprint arXiv:1211.3141*, 2012.

[17] ALFRPED RENYI. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.

[18] RS Varma. Generalizations of rényis entropy of order α. *J. Math. Sci*, 1(7):34–48, 1966.

[19] JN Kapur. Generalized entropy of order α and type β. In *The Math. Seminar*, volume 4, pages 78–82, 1967.

[20] Bhu Dev Sharma and Inder Jeet Taneja. Three generalized non-additive measures of entropy. *Elektronische Informationsverarbeitung und Kybernetik*, 13(7/8):419–433, 1977.

[21] CF Picard. Weighted probabilistic information measures. *Journal of Combinatorics, Information and System Sciences*, 4:343–356, 1979.

[22] G Rubiano. Topología general. *Universidad Nacional de Colombia, Bogotá*, 2002.

[23] Klaus Janich. Topology. undergraduate texts in mathematics, 1984.

[24] Eberhard Zeidler. *Nonlinear functional analysis and its applications*, volume 2. Springer, 1990.

[25] Walter Rudin. *Real and complex analysis (3rd)*. New York: McGraw-Hill Inc, 1986.

[26] Thomas Jech. *Set theory*. Springer Science, 2013.

[27] Rafael C Gonzalez, Richard E Woods, and Steven L Eddins. *Digital image processing using MATLAB*, volume 2. Gatesmark Publishing Knoxville, 2009.

[28] Prasanna Sahoo, Carrye Wilkins, and Jerry Yeager. Threshold selection using renyi's entropy. *Pattern recognition*, 30(1):71–84, 1997.

[29] M Portes de Albuquerque, IA Esquef, and AR Gesualdi Mello. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, 25(9):1059–1065, 2004.