"Vasile Alecsandri" University of Bacău
Faculty of Sciences
Scientific Studies and Research
Series Mathematics and Informatics
Vol. 19 (2009), No. 2, 257 - 264

# ANALYSIS OF ENTROPY MEASURES

ANGEL GARRIDO

**Abstract.** Our paper to develop some useful analytical tools for the foundations of Uncertainty Measures. Because we need to obtain new ways to model adequate conditions or restrictions, constructed from vague pieces of information. For this, it is necessary to classify more efficiently the distinct types of measures; in particular, the fuzzy measures.

Now, we complete this study by the analysis on Entropy and other Measures of Uncertainty, with their relationships. So, we attempt to go on, advancing by this paper.

## 1. Introduction

Kaufmann (1975) introduce the index of fuzziness as a normal distance.

Yager (1979), and Higashi and Klir (1983), shows the entropy measure as the difference between two fuzzy sets, a fuzzy set and its complementary, which is also a fuzzy set.

Taking the Entropy concept, we attempt to measure the fuzziness, that is, the degree of fuzziness for each element $A \in \wp$.

**Keywords and phrases:** ultrametric space, spherically complete, fixed point, implicit relation.
**2000 Mathematics Subject Classification**: 94A15, 94A17,37M25.

It can be designed as the function

$$H{:}\wp \rightarrow [0{,}1]$$

Verifying that

I) If $A$ is a *crisp set,* then $H(A) = 0$.

II) If $H(x) = 1/2$, for each $x$ in $A$, then $H(A)$ *is maximal* (i.e. the maximal uncertainty state).

III) If $A$ *is less fuzzified than* $B$, then $H(A) \leq H(B)$.

IV) $H(A) = H(U \setminus A)$.

Given a discrete random variable, X, with associated probability distribution P(x), we will define the *Entropy (H)* of $X$ as

$$H(x) \equiv -\sum_{x \in X} p(x) \ln p(x) = \sum_{x \in X} p(x) \ln (1/p(x)) = E\left[\ln \frac{1}{p(x)}\right]$$

Such H is a measure of the quantity of information that we receipt, when is sent towards us.

The logarithmic base will be arbitrary.

If $b = 2$, it is measured in *bits.*

If $b = 10$, it will be in *dits.*

And if $b = e$, in *nats.*

Given two random discrete variables, X and Y, their *Joint Entropy* is given by

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y) =$$
$$= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 [1 - p(x,y)] = -E[\log_2 p(x,y)]$$

The *Conditional Entropy* of the random variable Y, given the random variable X, will be introduced by

$$H(Y/X) = \sum_{x \in X} p(x) H(Y/X = x) =$$
$$-\sum_{x \in X} p(x) \sum_{y \in Y} p(y/x) \log_2 p(x,y) =$$
$$= \sum_{x \in X} p(x) \sum_{y \in Y} p(y/x) \log_2 [1 - p(x,y)]$$

On an infinite domain, it is possible to generalize the entropy.

We found the *Chain Rule,* involving the symmetry of this $H$ function

$$H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y) = H(Y,X)$$

It is also possible to prove that when X and Y are mutually equivalent random variables, then

$$H(X,Y) = H(X) + H(Y)$$

With the *Corollary,*

$$H(X,Y/Z) = H(X/Z) + H(Y/X,Z)$$

*Mutual Info between X and Y.* It will be denoted by *I (X; Y).*

But previously, we will define the *Relative or Differential Entropy.* It will be also called *Kullback-Leibler* (1951) *"distance"* (pseudo-distance, indeed), *D,* or *divergence K-L.*

Given two probability distributions, $p$ and $q$, it will be defined by

$$D(p \shortparallel q) = \sum_{x \in X} p(x) \log_2 \left( \frac{p(x)}{q(x)} \right) = E_{p(x)} \left[ \log_2 \frac{p(x)}{q(x)} \right]$$

Some essential properties of $D$ will be

I) $D(p \shortparallel q) \geq 0$

II) $D(p \shortparallel q) \geq 0 \; if \; and \; only \; if \; p(x) = q(x), \; \forall x$

III) In general, $D(p \shortparallel q) \neq D(q \shortparallel p)$

Therefore, it does not symmetrical. Neither verifies the triangular inequality. So, it is not really a metric.

It report us the measure of inefficiency when supposing $q$ as the correct distribution, being so indeed $p$.

The *mutual info of X on Y* is the measure of the info which X has on Y. Denoted by *I (X; Y).*

If we write instead *I (Y; X),* we have the info which $Y$ posess on *X.* But they give us the same value; so, it is *a symmetrical measure.*

The *relationship between mutual info and entropy* is

$$I(X;Y) = H(X) - H(X/Y) = H(Y) - H(Y/X) = I(Y;X)$$

And also

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Therefore,

$$I(X;X) = H(X)$$
$$I(Y;Y) = H(Y)$$

In general, conditioning a random variable on another, we reduce the uncertainty of the last variable

$$H\left(Y/X\right) \leq H\left(X\right)$$
$$H\left(X/Y\right) \leq H\left(Y\right)$$

It is possible to generalise the *Chain Rule for n variables*

$$H\left(X_1, H_2, ..., X_n\right) = \sum_{i=1}^{n} H\left(X_i / X_{i-1}, X_{i-2}, ..., X_1\right)$$

And therefore, in the conditional case

$$H\left(X_1, H_2, ..., X_n / Y\right) = \sum_{i=1}^{n} H\left(X_i / X_{i-1}, X_{i-2}, ..., X_1, Y\right)$$

*Generalizing the K-L divergence.*
In the *discrete case,* we have

$$D\left(P \shortparallel Q\right) = \sum_{i} p\left(i\right) \log_2 \left(\frac{P(i)}{Q(i)}\right)$$

Whereas in the *continuum case*

$$D\left(P \shortparallel Q\right) = \int_{-\infty}^{+\infty} p\left(x\right) \log_2 \left(\frac{p(x)}{q(x)}\right) dx$$

Being $p$ and $q$ the density functions corresponding to both, $P$ and $Q$ distributions.

Let

$$dP = p d\mu$$
$$dQ = q d\mu$$

be two probability measures, on the set X, such that they are absolutely continuous with respect to the measure.

Then, we define the *divergence of Kullback-Leibler, or K-L* (if such integral exist) as

$$D\left(P \shortparallel Q\right) = \int_{X} p \log \left(\frac{p}{q}\right) d\mu$$

Where

$$\frac{p}{q} = \frac{dP}{dQ}$$

is the *Radon-Nikodym derivative* of $P$ with respect to $Q$. Then, the final expression should be independent of measure $\mu$.

*Some interesting measures of divergence.* For instance, we have the *symmetrized distance*

$$D\left(P \shortparallel Q\right)+D\left(Q \shortparallel P\right)$$

It will be very useful, for instance, in Feature Selection, into Classification Problems.

An alternative distance is the $\lambda-\mathrm{div}$ *(lambda divergence),*

$$D_{\lambda}\left(P \shortparallel Q\right)=$$
$$\lambda D\left[P \shortparallel \lambda P+(1-\lambda) Q\right]+(1-\lambda) D\left[Q \shortparallel \lambda P+(1-\lambda) Q\right]$$

This signifies the gaining expectation of info about that X is obtained from $P$ or $Q$, with respective probabilities $p$ and $q$.

In particular, when $\lambda=1/2$, we found the *Jensen-Shannon divergence*

$$D_{JS}\left(P \shortparallel Q\right)=\tfrac{1}{2} D\left(P \shortparallel M\right)+\tfrac{1}{2} D\left(Q \shortparallel M\right)$$

Where $M$ is the promediate value of probability distributions $P$ and $Q$.

This *divergence of Jensen-Shannon* can be interpreted as the capability of a noisy channel of info with two entries and giving as output the probability distributions $P$ and $Q$.

More generalized, from the Shannon Entropy measure, we can found the *Rényi Entropy,* or Entropy due to Alfred Rényi.

Let be a random sample, $\{x_i\}_{i=1}^{n}$, with probabilities $\{p_i\}_{i=1}^{n}$.

We define the *Rényi´s Entropy* as

$$H_{\alpha}(X)=\tfrac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} p_i^{\alpha}\right)$$

If they are equal all the above probabilities, then

$$H_{\alpha}(X)=\log n, \ \forall \alpha$$

The entropies, as functions of $\alpha$, are weakly decreasing.

So, for instance,

$$H_0(X) \geq H_1(X) \geq H_2(X) \geq \ldots \geq H_{\infty}(X)$$

A particular case should be the *Hartley´s entropy,*

$$If\ \alpha=0, then\ H_0(X)=\log\ n\,(\log\left[card\,(X)\right])$$

There exists these relation between entropies

$$H_{\infty}<H_2<2 H_{\infty}$$

Furthermore, the *Generalized Divergence of Rényi,* of order $\alpha$, of a distribution *Q,* relative to *P,* the "authentic", will be

$$D_\alpha \left( P \shortparallel Q \right) = \frac{1}{\alpha - 1} \log \left( \sum_{i=1}^{n} \frac{p_i^\alpha}{q_i^{i-1}} \right) = \frac{1}{\alpha - 1} \log \left( \sum_{i=1}^{n} p_i^\alpha q_i^{1-i} \right)$$

So, we have

$$D_\alpha \left( P \shortparallel Q \right) \geq 0, \forall P, Q$$

## 2. Generalization

We may see now the *Entropy Measures* in a more generalized version.

Departing of a *t-norm, T,* a *t-conorm, S,* and the negation, N, it will be possible to introduce the Entropy Measure in another way.

Let

$$H : \wp \rightarrow [0, 1]$$

$$H\left(m\right) = k \; S \left\{ T \left( m \left( x \right) \right), N \left( m \left( x \right) \right) \right\}_{x \in U}$$

Where k is a *constant of normalization,* and $m$ a *fuzzy set* on *U,* wich depends of *T, S* and *N.*

We have the subsequent *results,*

1) If $m$ is a *crisp set,* then *H (m) = 0.*

2) When $T$ is in the minimum in the family of the t-norm product, then *H (m) = 0* if and only if *m is a crisp set.*

3) If $A$ is *less fuzzified than B,* then $H(A) \leq H(B)$.

4) *H (A) = H (U \ A).*

5) *H (m) = H (c(m)).*

6) Let $T$ as in 2). Then, if $B \leq_S A$, then $H\left(B\right) \leq H\left(A\right)$.

In this manner, we found that the Entropy Measure verifies the properties signaled by De Luca and Termini (1972).

On an *infinite domain,* it is possible to generalize the entropy by

$$H\left(m\right) = k \int_{x \in X} \left\{ T \left( m \left( x \right), N \left( m \left( x \right) \right) \right) \right\} \; dx$$

## 3. Conclusion

With this new approximation to fuzzy measures and their classification, we hope to contribute in the advance through the field of Uncertainty Measures, for to give an example of application. And so, advancing through the Approximate Reasoning. It will be also useful to work in different fields, such as Fuzzy Inference, Fuzzy Optimization, and so on.

## References

[1] Dubois and Prade (eds.), **Fundamentals of Fuzzy Sets,** Kluwer Acad. Publ., Dordrecht, Holland, 2000.

[2] Garrido, **"Additivity and Monotonicity in Fuzzy Measures",** plenary talk in ICMI 45, at Bacau University. Studii si Cercetare Stiintifice - Seria Matematica (SCSSM), Vol. No. 16, Supplement (2006): 445-458, Bacau.

[3] Garrido, **"Classifying Fuzzy Measures",** Acta Universitatis Apulensis (AUA) journal, Alba Iulia University, ICTAMI, Vol. No. 15 (2007): 445-458.

[4] Grabisch, Murofushi and Sugeno, **"Fuzzy measure of fuzzy events defined by fuzzy integrals",** Fuzzy Sets and Systems 50 (1992): 293-313, .

[5] Halpern, **Reasoning about Uncertainty,** MIT Press, 2005.

[6] Kullback, **Information Theory and Statistics,** Wiley, 1959.

[7] Rényi, **Selected Papers.** Akademiai Kiado, Budapest, 1976.

[8] Shannon, **A Mathematical Theory of Communication,** Illinois Press, 1963.

Angel Garrido
Department of Fundamental Mathematics
National University of Distance Learning (UNED)
Senda del Rey, 9. 28040-Madrid, Spain
email: *agarrido@mat.uned.es*