Tsalli's Entropy, Renyi's Entropy and its Applications

Paridhi Prasad
Nemani Harshavardhan
Sreeja Guduri
Bodgam Rohan Reddy

December 3, 2023

Introduction

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

- Entropy was first introduced in thermodynamics and has been developed into various forms from Boltzmann in statistical mechanics to Von Neumann in quantum mechanics.
- It was in late 1940s when Shannon has developed his own definition of entropy and defined it in the filed of information theory. He defined it as information entropy and this was the first time entropy was used in mathematics. Generally konwn as Shannon's entropy.
- In 1988, Tsallis developed the concept of his version of entropy named tsalli's entropy which was a generalisation of Shannon's entropy. Even though his version of entropy was debatable at first, it had seen uses in various complex systems like hydrological, hydraulic processes.

Definition of Tsallis Entropy

For a discrete random variable X(=x_i,i=1,2,...,n) with probability distribution P(=p_i,i=1,2,...,n), tsallis entropy is defined as

$$S = k * \frac{1 - \sum_{i=1}^{n} p_i^m}{m-1} = \frac{k}{m-1} * \sum_{i=1}^{n} (p_i - p_i^m)$$

Generally, the value of k=1 in most cases. Similarly, for a continuous random variable X with probability density function p(x), tsallis entropy is defined as

$$s = \frac{1 - \int_{-\infty}^{+\infty} p_i^m}{m-1}$$

TSALLIS ENTROPY AND SHANNON ENTROPY

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

For a discrete random variable X(=x_i,i=1,2,...,n) with probability distribution P(=p_i,i=1,2,...,n), shannon entropy is defined as

$$H(X) = -\sum_{i=1}^{n} p_i * log(p_i)$$

As m $\rightarrow$ 1, tsallis entropy converges to shannon entropy which means tsallis entropy is a generalization of shannon entropy.

$$\lim_{m\rightarrow 1} S = \lim_{m\rightarrow 1} \frac{1-\sum_{i=1}^{n} p_i^m}{m-1} = -\sum_{i=1}^{n} p_i * log(p_i) = H(X)$$

The statement also holds true for tsallis entropy of continuous variable and differential entropy.

Maximum value: Tsallis entropy takes maximum value for all values of m when the outcomes of the random variable are equally likely. When $p_i = 1/n \ \forall i = 1, 2, \ldots n$, S takes maximum value.

Additivity: If A and B are two random variables,

$$S(A, B) = S(A) + S(B/A) + (1 - m)S(A)S(B/A)$$

and if A and B are independent,

$$S(A, B) = S(A) + S(B) + (1 - m)S(A)S(B)$$

Properties of Tsallis Entropy

Concavity: Tsalli's entropy is either convex or concave depending on the value of m. A function is called convex function if f"(x) $\geq$ 0 $\forall$ x and concave if f"(x) ¡ 0 $\forall$ x. Using Jensen's inequality, we can find out whether Tsallis entropy is convex or concave.

When m $>$ 0: Tsallis entropy is a concave function.
When m $\leq$ 0: Tsallis entropy is a convex function.

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Shannon-Khinchin Axioms

1. For any given distribution $(p_1, p_2, p_3, ..., p_n)$, the function $(p_1, p_2, p_3, ..., p_n)$ takes maximum value only when $p_i = \frac{1}{n}$

$$S(p_1, p_2, p_3, ..., p_n) \leq S(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$$

2. If $p_{ij} \geq 0$, $p_i = \sum_{j=1}^{m_i} p_i j, \forall i = 1, ...n, \forall j = 1, ..m_i$, then

$$S(p_{11}, ...p_{nm_n} = S(p_1, p_2, ..., p_n) + \sum_{i=1}^{n} p_i * S(\frac{p_{i1}}{p_i}, ..., \frac{p_{im_i}}{p_i})$$

3. $S(p_1, p_2, ..., p_n, 0)$ = $S(p_1, ..., p_n)$

Let S($p_1, p_2, ..., p_n$) be a function for any integer n and values p_i such that $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$. If for any n, we have the function to be continuous with respect to all the Shannon-Khinchin arguments and satisfies them, then S can be defined as:

$$S(p_1, p_2, ..., p_n) = \frac{1 - \sum_{i=1}^{n} p_i^m}{\phi(m)}$$

where m $\in R^+$ and $\phi(m)$ satifies some properties.

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Properties satisfies by $\phi(m)$

1. $\phi(m)$ is continuous and $\phi(m) * (m-1) > 0$
2. $\lim_{m \to 1} \phi(m) = \phi(1) = 0$
3. The function $\phi(m)$ must be differentiable in the interval (a,1) $\bigcup$ (1,b)
3. There exists a constant k $>$ 0 such that $\lim_{m \to 1} \frac{d\phi(m)}{dm} = \frac{1}{k}$

Definition of Renyi's Entropy

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

* The first instance of a generalised entropy being used in information theory is attributed to Alfred Renyi.
* Renyi's entropy is a parametric family of information measures that depend on the parameter α. For special values of α, the notion of Renyi's Entropy reproduces various special entropies.
* For a random variable X

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{N} P(X = x_i)^\alpha \right)$$

Properties

* $H_\alpha(X)$ is non-negative: $H_\alpha(X) \geq 0$
* $H_\alpha(X)$ is additive by construction.
* H_α can be defined for continuous random variables as

$$H_\alpha(X) = \lim_{n \to \infty} (I_\alpha(P_n) - \log n) = \frac{1}{1-\alpha} \log \int p^\alpha(x) dx$$

specifcally, for quadratic entropy ($\alpha = 2$):

$$H_2(X) = -\log \int p^2(x) dx$$

VARYING THE PARAMETER α

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

* **Max-Entropy/Hartley Entropy:** For $\alpha = 0$

$$H_\alpha(X) = \log(N)$$

* **Collision Entropy:** Indicates Probability of two Random Variables taking same value, for $\alpha = 2$,

$$H_\alpha(X) = -\log\left(\sum_{i=1}^{N} P(X = x_i)^2\right)$$

* **Shannon Entropy:** As $\alpha \to 1$:

$\lim_{\alpha\to1} H_\alpha(X) = \lim_{\alpha\to1} \frac{1}{1-\alpha} \log \sum_{k=1}^{N} p_k^\alpha = \frac{\lim_{\alpha\to1} \sum_{k=1}^{N} \log p_k p_k^\alpha / \sum_{k=1}^{N}}{\lim_{\alpha\to1} -1} = H_S(X)$

(using L'Hospitals rule)

* **Min-Entropy:** As $\alpha \to \infty$, we get the min-entropy which is characterised by the negative logarithm of probability of the most likely outcome.

Introduction to Uniqueness Theorem

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

* The uniqueness theorem aims to demonstrate that a function acting on a given probability distribution, while following the Shannon-Khinchin axioms and exhibiting additivity, will take the form of Renyi's entropy.
* For this proof, we use the generalised entropy form that was given by Daroczy in 1964.
  Given a probability distribution $P = (p_1, ..., p_n)$ with $\sum_{k=1}^{N} p_i \leq 1$, the entropy of P is given by

$$H(P) = G^{-1}\left(\frac{\sum_{i=1}^{N} f(p_i)G(-\log_2 p_i)}{\sum_{i=1}^{N} f(p_i)}\right)$$

Theorem 1 (Shannon-Khinchin Axioms). *Let P be the set of probability mass distribuitions $\{p_1, p_2, ...p_N\}$ for all $N \geq 2$. H is an entropic form.*

1. ***Continuity:*** *$H(p_1, p_2, ..., p_N)$ depends continuously on all variables for each n.*
2. ***Maximality:*** *For all n,*

$$H(p_1, p_2, ..., p_N) \leq H(\frac{1}{N}, ..., \frac{1}{N}).$$

3. ***Expansibility:*** *For all N and $1 \leq i \leq N$,*

$$H(0, p_1, ..., p_N) = H(p_1, ..., p_i, 0, p_{i+1}, ..., p_N) = H(p_1, ..., p_i, p_{i+1}, ..., p_N).$$

4. ***Separability:*** *(or Strong Additivity) For all N,M,*

$$H(p_{11}, ..., p_{1M}, p_{21}, ..., p_{2M}, ..., p_{N1}, ..., p_{NM})$$

The axioms that we consider are:

A.1. G is a strictly monotonic and continuous function.

A.2. f is a positive valued and bounded function defined in [0,1] such that

$$f(xy) = f(x)f(y)$$

A.3. $H(P * Q) = H(P) + H(Q)$, where for $P = (p_1, \dots, p_N)$ and $Q = (q_1, \dots, q_M)$,

$$P * Q = (p_1 q_1, \dots, p_1 q_M, \dots, p_N q_1, \dots, p_N q_M)$$

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

* We know,

$$H(p) = -\log_2 p$$

Using A.3, we have

$$H(p_1 q, p_2 q, ..., p_N q) = H(p_1, p_2, ...p_N) + H(q)$$

$$\Rightarrow G^{-1}\left(\frac{\sum_{i=1}^{N} f(p_i q) G(-\log_2 p_i q)}{\sum_{i=1}^{N} f(p_i q)}\right) = G^{-1}\left(\frac{\sum_{i=1}^{N} f(p_i) G(-\log_2 p_i)}{\sum_{i=1}^{N} f(p_i)}\right) - \log_2 q$$

Proof for Uniqueness Theorem

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

* Substituting $-\log p_i = x_i$ and $-\log_2 q = y$ and taking $G(x_i + y) = G_y(x_i)$,

$$\Rightarrow G_y^{-1}\left(\frac{\sum_{i=1}^{N} f(2^{-x_i})G_y(x_i)}{\sum_{i=1}^{N} f(2^{-x_i})}\right) = G^{-1}\left(\frac{\sum_{i=1}^{N} f(2^{-x_i})G(x_i)}{\sum_{i=1}^{N} f(2^{-x_i})}\right)$$

* For this equation to be strictly monotonic and continuous, in general

$$G_y(x) = G(x+y) = A(y)G(x) + B(y)$$

* For $G(x)$ strictly monotonic function gives only two solutions:

$$G(x) = ax + b, a \neq 0$$

and

$$G(x) = ae^{cx} + b, a \neq 0, c \neq 0$$

Proof for Uniqueness Theorem

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

* The function corresponding to the solution

$$G(x) = ae^{cx} + b$$

with $c = (1-\alpha)\log_e 2$ is:

$$H(P) = \frac{1}{1-\alpha}\log_2 \frac{\sum_{i=1}^{N} f(p_i)p_i^{\alpha-1}}{\sum_{i=1}^{N} f(p_i)}$$

which is the same as Renyi's entropy general form.

Generalization to Quantum

* Renyi's entropy nicely unifies different and isolated measures into an α parameterized entropy measure.
* Then, the quantum version of the classical Rényi entropy is given by

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho^\alpha)$$

where $\alpha \in (0,1) \cup (1,\infty)$ and $Tr()$ refers to the trace of the matrix.

* Just like in the classical case, the von Neumann entropy (which is the quantum analog of Shannon's entropy) is a limiting case of the Rényi entropy.

$$\lim_{\alpha \to 1} S_\alpha(\rho) = S(\rho)$$

Comparing Renyi's and Tsallis Entropy

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

* The following equations give a straightforward relation between Renyi's and Tsallis' Entropies:

$$T_\alpha = \frac{1}{1-\alpha}\left(e^{(1-\alpha)R_\alpha} - 1\right)$$

or,

$$R_\alpha = \frac{1}{1-\alpha}\ln\left(1 + (1-\alpha)T_\alpha\right)$$

—

* R_α is additive by construction while Tsallis entropy follows the relation of:

$$T_\alpha(A,B) = T_\alpha(A) + T_\alpha(B/A) + (1-m)T_\alpha(A)T_\alpha(B/A)$$

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

- Estimating information in natural data such as EEG data is fairly difficult since there isn't a relationship or encoding process between the physical data and encoded data. (also known as the Encoding problem)
- We also know that the physical instantiation of any information event, that is, the physical occurrence of a symbol of information, must begin and end at a discontinuity or critical point (maximum, minimum, or saddle point) in the data including EEG waveform events.
- The variance within of an EEG waveform event is directly proportional to its probability of occurrence

Applications in Natural Information Theory

Using the points of discontinuity, we can easily break the waveform into disjoint events and bin them into categories.

Since the variance is directly proportional to the probability of occurrence, we can count frequency and compute variances within the ESG waveform. Using this, we can derive a linear estimator of Tsallis's entropy functional.

$$\widehat{H}_T(Y) = 1 - \sum_{i=1}^{N} \widetilde{p}(\mathbf{y}_i)\widehat{p}(\mathbf{y}_i).$$

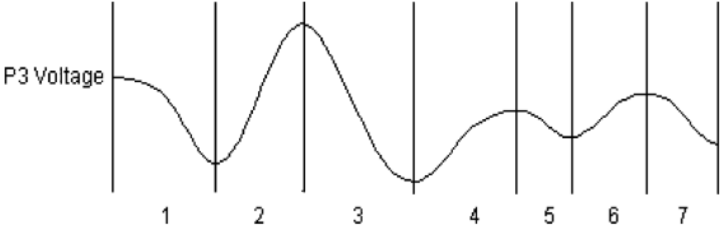
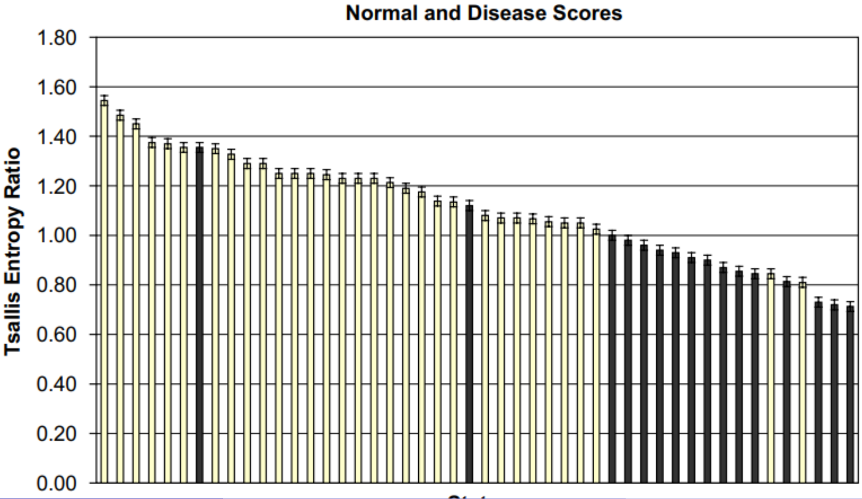


Fig. 2. An EEG waveform binned at critical points.

Applications in Natural Information Theory

We can use this estimator to predict ADRD (Alzheimer's and Related Diseases), as well as monitor its treatment. This approach has 91% accuracy, which shows that Tsallis's estimators are an efficient way to predict data in waveform formats without actually decoding.

Applications in Natural Information Theory



INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

The fractal dimension of weighted complex networks is obtained as follows,

$$d_{fra} = -\lim_{s \to 0} \frac{\ln N(s)}{\ln s}$$

where $N(s)$ is the number of boxes needed to cover the whole networks when the box size equals to s.
The local dimension is obtained as follows,

$$d_{i\_loc} \approx r \frac{n_i(r)}{N_i(r)}$$

where $n_i(r)$ is the number of nodes whose shortest distance from center node i equal to radius $r.N_i(r)$ is the number of nodes within the radius r (the distance can be less or equal to the radius r).

Measuring Complexity

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

The degree structure entropy is wildly used in the real-world networks' structure complexity measuring, and it is based on the Shannon entropy and the degree distribution, which is detailed shows as below,

$$E_{\text{deg}} = -\sum_{i=1}^{N} p_i \log(p_i)$$

where N is the total number of the nodes, and p_i degree dist of i,

$$p_i = \frac{S(i)}{\sum_{i=1}^{N} S(i)}$$

$$S(i) = \sum_{j=1}^{N} w_{ij}$$

where w_{ij} is the weight of edge connected with node i and node j.

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Because the degree structure entropy focuses on the local information, the betweenness structure entropy is proposed recently. It is based on the Shannon entropy and the betweenness distribution, which can be shown as follows,

$$E_{\text{bet}} = -\sum_{i=1}^{N} p'_i \log\left(p'_i\right)$$

where N represents the total number of nodes, and p'_i represents the betweenness distribution [74] of node i which can be detailed shown as follows,

$$p'_i = \frac{\sum_{s \neq t, s \neq i, t \neq i} L_{st}(i)}{\sum_{s \neq t} L_{st}}$$

where L_{st} is the total number of shortest paths between node s and node t, $L_{st}(i)$ is the number of shortest paths between node s and node t which pass through node i.

Applications in Networks

The proposed structure entropy is based on Tsallis entropy and is defined as follow,

$$T_{lf} = k \sum_{i=1}^{N} \frac{p_i^{d_{fra}} - p_i}{1 - d_{fra}}$$

where T_{lf} represents this proposed structure entropy of complex network, k equals to constant 1 in this method, p_i is related to the local dimension d_{i_loc} of node i

$$p_i = \frac{d_{i\_loc}}{\sum_{i=1}^{N} d_{i\_loc}}$$

The fractal dimension can reveal the whole structure property of a complex network, and the local dimension reveals the structure property based on the center node.

Image Thresholding

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Image segmentation is the process of getting objects and backgrounds in pictures that might not be cleanly extractable using CV-based approaches.

Tsallis and Renyi entropy is also considered as a generalization of Shannon entropy and explored in the field of image thresholding due to the desirable property of nonadditive information content in some image classes.

The probability distributions of the object and background classes, A_1 and A_2, are given by

$$A_1 : \frac{p_0}{p(A_1)}, \frac{p_1}{p(A_1)}, \ldots, \frac{p_t}{p(A_1)}$$

and

$$A_2 : \frac{p_{t+1}}{p(A_2)}, \frac{p_{t+2}}{p(A_2)}, \ldots, \frac{p_{255}}{p(A_2)}$$

where

$$p(A_1) = \sum_{i=0}^{t} p_i, p(A_2) = \sum_{i=t+1}^{255} p_i \text{ and } p(A_1) + p(A_2) = 1$$

Tsallis Entropy

$$S_q^{A_1}(t) = \frac{1 - \sum_{i=1}^{t} \left(\frac{p_i}{p^{A_1}}\right)^q}{q-1}$$

$$S_q^{A_2}(t) = \frac{1 - \sum_{i=t+1}^{k} \left(\frac{p_i}{p^{A_2}}\right)^q}{q-1}$$

We maximize the information measure between the two classes (object and background). When $S_q(t)$ is maximized, the luminance level t is considered to be the optimum threshold value.

$$t_{\text{opt}} = \operatorname{argmax} \left[S_q^{A_1}(t) + S_q^{A_2}(t) + (1-q) \cdot S_q^{A_1}(t) \cdot S_q^{A_2}(t) \right]$$

Renyi Entropy

$$H^{\alpha}_{A_1}(t) = \frac{1}{1-\alpha} \ln \sum_{i=0}^{t} \left(\frac{p_i}{p(A_1)} \right)^{\alpha}$$

and

$$H^{\alpha}_{A_2}(t) = \frac{1}{1-\alpha} \ln \sum_{i=t+1}^{255} \left(\frac{p_i}{p(A_2)} \right)^{\alpha}$$

Assuming $t^*(\alpha)$ is the optimum threshold value,

$$t^*(\alpha) = \operatorname*{Arg\,max}_{t \in G} \{H_{A_1}(t) + H_{A_2}(t)\}.$$

Renyi Entropy

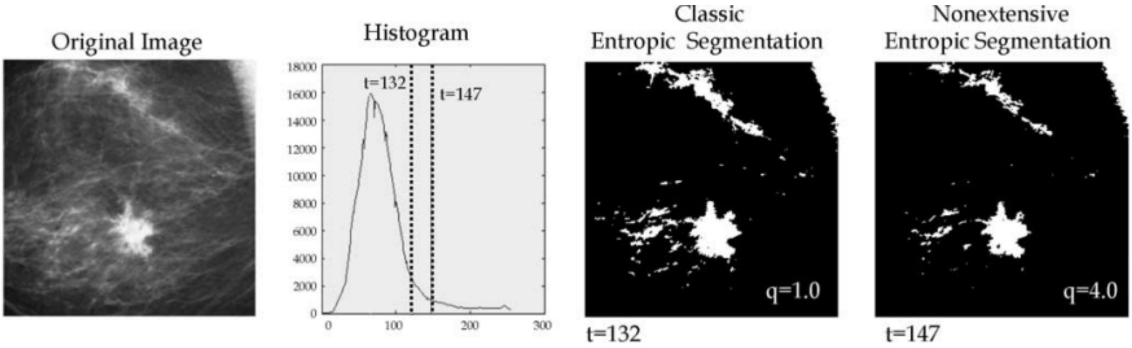


Figure: Example of entropic segmentation for mammography image with an inhomogeneous spatial noise. Two image segmentation results are presented for q = 1.0 (classic entropic segmentation) and q = 4:0.

Applications of Renyi in Finance

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

The rate of return expected by such an investor is equal to the relative entropy between the investor's believed probabilities and the official odds. The Rényi divergence of order α of b from m is:

$$D_\alpha(b\|m) \stackrel{\text{def}}{=} \frac{1}{\alpha-1}\ln\int b(x)\left(\frac{b(x)}{m(x)}\right)^{\alpha-1} dx$$

The expected profit rate is $= \frac{1}{R}D_1(b\|m) + \left(\frac{R-1}{R}\right)D_{1/R}(b\|m)$ Where b is the buyer's odds regarding an event, m is the market's odds, and R is the risk appetite of a buyer.

CONTRIBUTIONS

Rohan: Introduction to Tsallis entropy, properties, relation with shannon entropy and uniqueness theorem.

Paridhi: Introduction to Renyi's entropy and properties, Uniqueness Theorem for Renyi's Entropy

Sreeja: Uniqueness Theorem for Renyi's Entropy, Generalisation of Renyi's Entropy to Quantum

Harshavardhan: Applications

REFERENCES

Vijay P. Singh,"Introduction to Tsallis Entropy Theory in Water Engineering",2016.

Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. J Stat Phys 52, 479–487 (1988).

K. Tadaki, "The Tsallis entropy and the Shannon entropy of a universal probability," 2008 IEEE International Symposium on Information Theory, Toronto, ON, Canada, 2008, pp. 2111-2115, doi: 10.1109/ISIT.2008.4595362.

H. Suyari, "Generalization of Shannon-Khinchin axioms to nonextensive systems and the uniqueness theorem for the nonextensive entropy," in IEEE Transactions on Information Theory, vol. 50, no. 8, pp. 1783-1787, Aug. 2004, doi: 10.1109/TIT.2004.831749.

References

T. Maszczyk, W. Duch, ”Comparison of Shannon, Renyi and Tsallis Entropy Used in Decision Trees”

B. Sharma, ”Uniqueness Theorems of Generalized Entropy and Information”

M. Muller-Lennert et. al., ”On Quantum Renyi Entropies: A New Generalization and some Properties”

J. Amigo. ”A Brief Review of Generalized Entropies”

The End

Questions? Comments?