Entropy and Information - Project
Tsallis and Renyi Entropies and their Applications

Paridhi Prasad (2021102008), Nemani Harshavardhan (2020111022), Sreeja Guduri (2021102007), Bodgam Rohan Reddy (2020102039)

December 3, 2023

1 Introduction

Entropy is a quantity first introduced in Thermodynamics, and has subsequently been developed into various forms, ranging from Boltzmann Entropy in Statistical Mechanics to Von Neumann Entropy in Quantum Mechanics.

In the Information Theory context, Entropy was defined by Claude Shannon in the 1940s, as Information Entropy or Shannon Entropy. This was the first time Entropy was defined in the mathematical context.

Renyi and Tsallis Entropies are both Generalized Entropies, proposed by Constantino Tsallis and Alfred Renyi respectively. Generalized Entropies are non-negative functions defined over probability distributions that satisfy the first three Shannon-Khinchin axioms.

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,*

$$H(p_{11},...,p_{1M},p_{21}...,p_{NM}) = H(p_1,p_2,...,p_N) + \sum_{i=1}^{N} p_i H(\frac{p_{i1}}{p_i},\frac{p_{i2}}{p_i},...,\frac{p_{iM}}{p_i})$$

where $p_i = \sum_{j=1}^{M} p_{ij}$

Let $(p_{11},...,p_{1M},p_{21},...,p_{2M},...,p_{N1},...,p_{NM})$ be the joint probability distribution of the random variables X and Y, with marginal distributions $\{p_i : 1 \leq i \leq N\}$ and $\{p_j = \sum_{i=1}^{N} p_{ij} : 1 \leq j \leq M\}$, respectively. Then, axiom 4 can be rewritten as

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

where $H(Y|X)$ is the entropy of Y conditioned over X.

A point of intrigue is that Renyi Entropy satisfies the fourth axiom as well, however Tsallis Entropy does not. Shannon Entropy can be approximated from both Renyi and Tsallis Entropy.

2 Tsallis Entropy

2.1 Definition

For a discrete random variable X($X = x_1, x_2, x_3, ..., x_n$) with probability distribution P($P = p_i, i = 1,2,3,.....,n$ and $\sum_{i=1}^{n} p_i = 1$), the tsallis entropy S_q is defined as,

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

where $k > 0$ (=1 in general cases) and $m \in R$.
And for the same discrete random variable X with probability distribution P, Shannon's entropy is defined as,

$$H = -\sum_{i=1}^{n} p_i * log(p_i) \tag{2}$$

For a continuous random variable X with probability density function p(x) (where $\int_{-\infty}^{+\infty} p(x) = 1$), tsallis entropy for X is defined as,

$$s = k * \frac{1 - \int_{-\infty}^{+\infty} p_i^m}{m-1} \tag{3}$$

and Shannon's entropy for a continuous variable or differential entropy is defined as,

$$h = -\int_{-\infty}^{+\infty} p_i * log(p_i) \tag{4}$$

Example 1: Let X be a random variable which takes two values and p, 1-p be the probability distribution of X.Plot Tsallis entropy versus probability p for values of m= -1,-0.5,0,0.5,1,2 for the random variable x

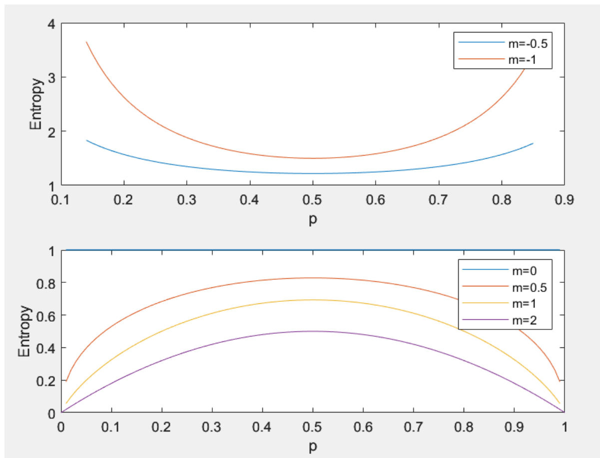

Figure 1: Tsallis entropy for a two values function for different values of m

2.2 Relation of Tsallis Entropy with Shannon's Entropy

Tsallis entropy is a generalised form of Shannon's entropy. When m→ 1, Tsallis entropy converges to Shannon entropy.

Proof : Tsallis entropy S = $\frac{1-\sum_{i=1}^{n} p_i^m}{m-1}$

$$\begin{aligned} \lim_{m\to 1} S = \lim_{m\to 1} \frac{1-\sum_{i=1}^{n} p_i^m}{m-1} \\ = \lim_{m\to 1} \frac{1-\sum_{i=1}^{n} p_i * p_i^{m-1}}{m-1} \\ = \lim_{m\to 1} \frac{1-\sum_{i=1}^{n} p_i * e^{m-1*log(p_i)}}{m-1} \end{aligned} \tag{5}$$

Applying L'Hôpital's rule which states that if for two functions f(x) and g(x) differentiable on an open interval I and for a point c in the interval, if $\lim_{x\to c}$ f(x) = 0 and $\lim_{x\to c}$ g(x) = 0, and g'(x) ≠ 0 for all x in I, then

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)} \tag{6}$$

In equation (3), f(x) $= 1 - \sum_{i=1}^{n} p_i * e^{(m-1)*log(p_i)}$ and g(x) = m-1, as m $\rightarrow$ 1, f(x) $= 1 - \sum_{i=1}^{n} p_i = 1 - 1 = 0$ and g(x) = 1 - 1 = 0. g'(x) = 1 $\neq 0 \; \forall$ m. Applying L'Hôpital's rule on equation (3), we get

$$
\begin{aligned}
\lim_{m \to 1} S &= \lim_{m \to 1} \frac{-\sum_{i=1}^{n} p_i * e^{m-1*log(p_i)} * log(p_i)}{1} \\
&= -\sum_{i=1}^{n} p_i * e^{1-1*log(p_i)} * log(p_i) \\
&= -\sum_{i=1}^{n} p_i * e^{0} * log(p_i) \\
&= -\sum_{i=1}^{n} p_i * log(p_i) = H(X)
\end{aligned}
$$

As m $\rightarrow$ 1, tsallis entropy of a continuous random variable converges to differential entropy.

Proof : Tsallis entropy for continuous random variable s = $\frac{1-\int_{-\infty}^{+\infty} p_i^m}{m-1}$

$$
\begin{aligned}
\lim_{m \to 1} s &= \lim_{m \to 1} \frac{1 - \int_{-\infty}^{+\infty} p_i^m}{m-1} \\
&= \lim_{m \to 1} \frac{1 - \int_{-\infty}^{+\infty} p_i * p_i^{m-1}}{m-1} \\
&= \lim_{m \to 1} \frac{1 - \int_{-\infty}^{+\infty} p_i * e^{m-1*log(p_i)}}{m-1} \qquad (7)
\end{aligned}
$$

In equation (7), f(x) $= 1 - \int_{-\infty}^{+\infty} p_i * e^{(m-1)*log(p_i)}$ and g(x) = m-1, as m $\rightarrow$ 1, f(x) $= 1 - \int_{-\infty}^{+\infty} p_i = 1 - 1 = 0$ and g(x) = 1 - 1 = 0. g'(x) = 1 $\neq 0 \; \forall$ m. Applying L'Hôpital's rule on equation (7), we get

$$
\lim_{m \to 1} s = \lim_{m \to 1} \frac{-\sum_{i=1}^{n} p_i * e^{m-1*log(p_i)} * log(p_i)}{1}
$$

From Leibniz integral rule, $\frac{d}{dx}(\int_a^b f(x,t)dt) = \int_a^b \frac{\partial}{\partial x} f(x,t)dt$

$$f'(x) = \frac{d}{dx}(f(x)) = \frac{d}{dx}(\int_{-\infty}^{+\infty} p_i * e^{(m-1)*log(p_i)})$$
$$f'(x) = \int_{-\infty}^{+\infty} (\frac{\partial}{\partial x} p_i * e^{(m-1)*log(p_i)}$$
$$f'(x) = \int_{-\infty}^{+\infty} p_i * e^{(m-1)*log(p_i)} * log(p_i)$$
$$\lim_{m \to 1} s = -\int_{-\infty}^{+\infty} p_i * e^{1-1*log(p_i)} * log(p_i)$$
$$= -\int_{-\infty}^{+\infty} p_i * e^0 * log(p_i)$$
$$= -\int_{-\infty}^{+\infty} p_i * log(p_i) = h(X)$$

2.3 Properties of Tsallis Entropy

2.3.1 Concavity

A function f(x) is called a concave function if its second derivative f"(x) ≥ 0 anywhere in the domain of X. And a function f(x) is a concave function when it is not convex. f(x) is concave if f"(x) ¡ 0 anywhere.

$$S = \frac{1 - \sum_{i=1}^{n} p_i^m}{m-1}$$
$$\frac{\partial S}{\partial p_i} = \frac{1}{m-1} * (-m * p_i^{m-1})$$
$$\frac{\partial^2 S}{\partial p_i^2} = -m * p_i^{m-2}$$

The concavity of Tsallis entropy depends on the value of m.

- If m > 0: $\frac{\partial^2 S}{\partial p_i^2} < 0$, and hence it is a concave.
- If m ≤ 0: $\frac{\partial^2 S}{\partial p_i^2} \geq 0$, and hence it is convex.

We can also figure out the concavity of Tsallis entropy using Jensen's inequality. Jensen's inequality states that a function f(x) defined on a random variable X will be a convex function if E[f(X)] $\geq$ f(E[X]).

2.3.2 Maximum value

Using lagrangian equation to maximize S subject to $\sum_i p_i$,

$$S = \frac{1 - \sum_{i=1}^{n} p_i^m}{m-1}$$
$$J = S + \lambda_0 * (\sum_i p_i - 1)$$
$$\frac{\partial J}{\partial p_i} = \frac{\partial S}{\partial p_i} + \lambda_0$$

Evalating $\frac{\partial J}{\partial p_i} = 0$ to find extremum.

$$\frac{\partial S}{\partial p_i} = \frac{1}{m-1} * (-m * p_i^{m-1})$$
$$\frac{\partial J}{\partial p_i} = \frac{1}{m-1} * (-m * p_i^{m-1}) + \lambda_0$$
$$\frac{\partial^2 S}{\partial p_i^2} = -m * p_i^{m-2}$$
$$\frac{\partial J}{\partial p_i} = 0$$
$$\frac{-m * p_i^{m-1}}{m-1} + \lambda_0 = 0$$
$$p_i = \frac{\lambda_o * (m-1)}{m}^{\frac{1}{m-1}}$$

For a given value of m, all p_i must be equal to attain either maximum or minimum value depending on whether the function is convex or concave.
Probability distribution for attaining maximum or minimum value $\left\{\frac{1}{N} . \frac{1}{N}, ..., \frac{1}{N}\right\}$ if X takes N unique values with probabilities p_i, i=1,2,...,n. To achieve this the constant λ_0 is used.

$$\lambda_0 = \frac{m}{m-1} * N^{1-m}$$

2.3.3 Additivity

For two independent random variable A and B,the additivity is defined as A $\bigcup$ B In case of Shannon's entropy, additive property is defined as

$$H(A, B) = H(A) + H(B) \tag{8}$$

In case of Tsallis entropy, additive property is defined as

$$S(A, B) = S(A) + S(B) + (1-q) * S(A) * S(B) \tag{9}$$

Proof: For two independent random variables A and B with probability distribution p_i for i=1,2,...N and q_j for j=1,2,...M, p(A + B) = p(A $\bigcup$ B) = p(A,B)

= p(A)*p(B)
p(A=i,B=j) = p(A=i)*p(B=j) = $p_i * q_j$

$$
\begin{aligned}
S(A) &= \frac{1-\sum_{i=1}^{N} p_i^m}{m-1} \\
S(B) &= \frac{1-\sum_{j=1}^{M} q_i^m}{m-1} \\
S(A,B) &= \frac{1-\sum_{i,j}^{N,M} [p(A,B)]^m}{m-1} \\
&= \frac{1-\sum_{i,j}^{N,M} (p_i * q_j)^m}{m-1} \\
\sum_{i,j}^{N,M} [p(A,B)]^m &= \sum_{i=1}^{N} (p_i)^m * \sum_{j=1}^{M} (q_j)^m
\end{aligned}
$$

Taking logarithm on both sides

$$
\begin{aligned}
log[\sum_{i,j}^{N,M} [p(A,B)]^m] &= log[\sum_{i=1}^{N} (p_i)^m] + log[\sum_{j=1}^{M} (q_j)^m] \\
log[\sum_{i=1}^{N} (p_i)^m] &= log(1 - \frac{(m-1)*[1-\sum_{i=1}^{N}(p_i)^m]}{m-1} \\
&= log[1-(m-1)S(A)]
\end{aligned}
$$

Similarly

$$
\begin{aligned}
log[\sum_{j=1}^{M} (q_j)^m] &= log[1-(m-1)S(B)] \\
log[\sum_{i,j}^{N,M} [p(A,B)]^m] &= log[1-(m-1)S(A,B)] \\
log[1-(m-1)S(A,B)] &= log[1-(m-1)S(A)] + log[1-(m-1)S(B)] \\
1-(m-1)*S(A,B) &= [1-(m-1)*S(A)]*[1-(m-1)*S(B)] \\
S(A,B) &= S(A)+S(B)+(1-m)*S(A)*S(B)
\end{aligned}
$$

Generalized Shannon Addivity Principle: For two random variables A and B,

$$S(A,B) = S(A) + S(B/A) + (1-m)*S(A)*S(B/A) \tag{10}$$

This can be proved using the equation,

$$P(A,B) = P(A) * P(B/A)$$

where in case of independence between A and B,

$$P(B/A) = P(B)$$

2.4 Uniqueness theorem of Tsallis Entropy

2.4.1 Shannon-Khinchin Axioms

1. Continuity: for any n $\in$ N and $p_1, p_2, ..., p_n$ be n different probabilities such that $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$, S($p_1, p_2, ...p_n$) is continuous for m $\in R^+$
2. Maximality: for any m $\in R^+$, $S(p_1, p_2, ..., p_n) \leq S(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$
3. Generalized Shannon Additivity: if $p_{ij} \geq 0, p_i = \sum_{j=1}^{M} p_{ij} \forall i = 1, 2, ...N and j = 1, ...M$, then

$$S(p_{11}, p_{12}..., p_{NM}) = S(p_1, p_2, ...p_N) + \sum_{i=1}^{N} p_i^m * S(\frac{p_{i1}}{p_i}, ...\frac{p_{iM}}{p_i}) \tag{11}$$

4. Expandability: $S(p_1, p_2, p_3, ...p_n, 0) = S(p_1, ..., p_n)$

Proof:
Axiom 1: Tsallis entropy is defined every for all value of m and the limit exists at m=1 which means S is continuous.
Axiom 2: Maximum value property of Tsallis entropy proved in section 2.3.2
Axiom 3: Let A,B are two dependent random variables with probability distributions p_i, i=1,...,N and p_j, j=1,....M such that P(A,B) = p_{ij}.

$$\begin{aligned}
S(A,B) &= S(p_{11}, p_{12}..., p_{NM}) \\
S(A) &= S(p_1, p_2, ...p_n) \\
S(A,B) &= \frac{1 - \sum_{i,j}^{N,M} [p(A,B)]^m}{m-1} \\
&= \frac{1 - \sum_{i=1}^{n} p_i^m + \sum_{i=1}^{n} p_i^m - \sum_{i,j}^{N,M} [p(A,B)]^m}{m-1} \\
S(A,B) &= \frac{1 - \sum_{i=1}^{N} p_i^m}{m-1} + \frac{\sum_{i=1}^{n} p_i^m - \sum_{i,j}^{N,M} [p_i * p_{ij}/p_i]^m}{m-1} \\
&= S(A) + \sum_{i=1}^{n} p_i^m * \frac{1 - \sum_{j}^{M} [p_{ij}/p_i]^m}{m-1} \\
S(p_{11}, p_{12}..., p_{NM}) &= S(p_1, p_2, ...p_N) + \sum_{i=1}^{N} p_i^m * S(\frac{p_{i1}}{p_i}, ...\frac{p_{iM}}{p_i})
\end{aligned}$$

Axiom 4:

$$\begin{aligned}
S(p_1,p_2,p_3,.....p_n,0) &= \frac{1-\sum_{i=1}^{n+1} p_i^m}{m-1} \\
&= \frac{1-p_1^m+p_2^m+.....+p_n^m+0^m}{m-1} \\
&= \frac{1-\sum_{i=1}^{n} p_i^m}{m-1} \\
&= S(p_1,p_2,p_3,....p_n)
\end{aligned}$$

2.4.2 Uniqueness Thoerem for Shannon's Entropy

Statement: Let $S(p_1,p_2,....,p_n)$ be a function defined for any integer n, then for all values $p_1,p_2,....,p_n$ such that $p_i \geq 0,(i=1,.....n)$ and $\sum_{i=1}^{n} p_k = 1$, if for any n, if it satisfies Shannon-Khinchin axioms, then

$$S(p_1,p_2,.....p_n) = \frac{1-\sum_{i=1}^{n} p_i^m}{\phi(m)} \tag{12}$$

where $q \in R^+$ and $\phi(m)$ follows the properties.

i) $\phi(m)$ is continuous and has same sign as (m-1), i.e.,

$$\phi(m)*(m-1) > 0 \qquad\qquad (m \neq 1)$$

ii) $\lim_{m\to 1}\phi(m) = \phi(1) = 0, \qquad \phi(m) \neq 0$ when $m \neq 1$

iii) there exists an interval (a,b) such that a¡1¡b and $\phi(m)$ is differentialble on the interval (a,1) $\bigcup$ (1,b)

iv) there exists a constant k > 0 such that, $\lim_{m\to 1}\frac{d\phi(m)}{dm} = \frac{1}{k}$

Proof:

Let for any i=1,.....r and j=1,.....s, if we take it such that $p_i = \frac{1}{r}\forall i = 1,...r$ and $p_j = \frac{1}{s}\forall j = 1,2,...,s$, and $p_ij = \frac{1}{rs}$, applying Shannon Khinchin's axiom (3), we get

$$S(\frac{1}{rs},.....\frac{1}{rs}) = S(\frac{1}{r},...,\frac{1}{r}) + \sum_{i=1}^{r}\frac{1}{r^m}*S(\frac{1}{s},....\frac{1}{s})$$

Let f(r) $= S(\frac{1}{r},...,\frac{1}{r})$, then

$$\begin{aligned}
f(rs) &= f(r) + \frac{r}{r^m}*S(s) \\
f(rs) &= f(r) + r^{1-m}*S(s)
\end{aligned}$$

Interchanging the variables, we get the same LHS but different RHS, equating both RHS after interchanging variables

$$\begin{aligned} f(r) + r^{1-m} * f(s) &= f(s) + s^{1-m} * f(r) \\ \frac{f(r)}{1 - r^{1-m}} &= \frac{f(s)}{1 - s^{1-m}} \end{aligned} \tag{13}$$

Equation (13) holds true for any r,s $\in$ N as the value of the ratio depends on m but not the variables r,s. Thus, we can say that there exists a function $\phi(m)$ such that,

$$f(n) = \frac{1 - n^{1-m}}{\phi(m)} \tag{14}$$

For all $p_i \in Q$, there exists non-negative integers z_i satisfying

$$p_i = \frac{z_i}{\sum_{k=1} n z_k} \tag{15}$$

And let p_{ij} be defined as

$$p_{ij} = \frac{1}{\sum_{k=1} n z_k}$$

for any i=1,....n and j==1,....,z_i, applying shannon-khinchin's axiom (3)

$$S(\frac{1}{\sum_{k=1} n z_k}, ..., \frac{1}{\sum_{k=1} n z_k}) = S(p_1....p_n) + \sum_{i=1}^{n} p_i^m * S(\frac{1}{z_i}, ...\frac{1}{z_i})$$

As f(n) = $S(\frac{1}{n}, ..., \frac{1}{n})$,

$$\begin{aligned} f(\sum_{k=1}^{n} z_k) &= S(p_1, p_2, ...p_n) + \sum_{i=1}^{n} p_i^m * f(z_i) \\ S(p_1, p_2, ...p_n) &= f(\sum_{k=1}^{n} z_k) - \sum_{i=1}^{n} p_i^m * f(z_i) \end{aligned} \tag{16}$$

Substituting eq. (14) in (16), we get

$$S(p_1, p_2, ...p_n) = \frac{1 - \sum_{i=1}^{n} p_i^m + \sum_{i=1}^{n} p_i^m * z_i^{1-m} - (\sum_{k=1}^{n} z_k)^{1-m}}{\phi(m)} \tag{17}$$

$$\begin{aligned} \sum_{i=1}^{n} p_i^m * z_i^{1-m} &= \sum_{i=1}^{n} (\frac{z_i}{\sum_{k=1}^{n} z_k})^m * z_i^{1-m} \\ &= \sum_{i=1}^{n} \frac{z_i}{(\sum_{k=1}^{n} z_k)^m} \\ &= (\sum_{k=1}^{n} z_k)^{1-m} \end{aligned} \tag{18}$$

Substituting in equation (17), we get

$$S(p_1, ...p_n) = \frac{1 - \sum_{i=1}^{n} p_i^m}{\phi(m)} \tag{19}$$

From Shannon-Kinchin axiom (1), we can prove that for any values $(p_1, p_2, ...p_n)$, the above equation is satisifed.
The numerator of equation (19) is a concave function when m¿1 and a convex function when 0¡m¡1 and is also symmetric with respect to values $p_1, ...p_n$. For the Shannon-Khinchin axiom (2) to be true, the function phi(x) must also follow that the value obtains maximality and is a concave function for a fixed m¿0. Hence $\phi(m) * (m - 1) > 0$ so that S is a concave function and obtains maximality We know that when $m \to 1$, S converges to uniqueness of shannon's entropy.

$$\lim_{m \to 1} \frac{1 - \sum_{i=1}^{n} p_i^m}{\phi(m)} = -\lambda * \sum_{i=1}^{n} p_i * lnp_i \tag{20}$$

And

$$\lim_{m \to 1} 1 - \sum_{i=1}^{n} p_i^m = 0 \tag{21}$$

Also, the numerator is differential with respect to m for any value of m.

$$\frac{d(numerator)}{dm} = -\sum_{i=1}^{n} p_i^m * ln(p_i) \tag{22}$$

When we substitute m=1, we get

$$\frac{d(numerator)}{dm}(m=1) = -\sum_{i=1}^{n} p_i * ln(p_i) \tag{23}$$

So, for the limit to be true, the value of $\phi(m) = 0$ when m=1. Hence (ii) is proved. Also, $d\phi(m)/dm$ must exist leading to $\phi(m)$ satisfying (iii). MOreover, $d\phi(m)/dm$ at m $\to$ 1, should be

$$\lim_{m\to 1} \frac{d\phi(m)}{dm} = \frac{- * \sum_{i=1}^{n} p_i * lnp_i}{-\lambda \sum_{i=1}^{n} p_i^m * ln(p_i)} = \frac{1}{\lambda} \tag{24}$$

Therefore, from the uniqueness theorem, if $S(p_1, p_2, ..., p_n)$ be a function defined for any integer n, then for all values $p_1, p_2, ..., p_n$ such that $p_i \geq 0, (i = 1, ...n)$ and $\sum_{i=1}^{n} p_k = 1$, if for any n, if it satisfies Shannon-Khinchin axioms, then

$$S(p_1, p_2, ...p_n) = \frac{1 - \sum_{i=1}^{n} p_i^m}{\phi(m)} \tag{25}$$

and the function $\phi(m)$ follows the properties (i,ii,iii,iv).

3 Renyi Entropy

The first instance of a generalised entropy being used in information theory is attributed to Renyi, who in the early 60's, introduced a new parametric form, that reduces to the classic Shannon Entropy for specific values of the parameter.

Alfred Rényi was trying to look for the most generic way of quantifying information while preserving the additivity of independent events and being compatible with the Shannon–Khinchin axioms.

We know that the Shannon entropy $H(X)$ is given by

$$H_S(X) = -\sum_i P(x_i) \cdot \log_2(P(x_i))$$

where $P(x_i)$ is the probability mass function of a random variable X.

This definition has an implicit assumption: It uses the linear average of the information function $I(p) = -\log_2(p)$.

If instead we use the general theory of means, then the average of any function $f(x)$ with an inverse $f^{-1}(x)$ can be given by

$$Average = f^{-1}(\sum_{i=1}^{N} p_k f(x_k))$$

Applying this definition to $I(p)$ we get

$$I(p) = f^{-1}(\sum_{i=1}^{N} p_k f(I_k))$$

For the postulate of addivity of independent events to be satisfied, we need

$$H(AB) = H(A) + H(B)$$

This constraint gives us two possibilities for $f(x)$:

$$f(x) = ex$$

and,

$$f(x) = e^{-2(1-\alpha)x}$$

Using $f(x) = cx$ gives us Shannon information and using the second form will gives us

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{N} P(X = x_i)^\alpha \right) \tag{26}$$

This gives a parametric family of information measures that are called Renyi's entropies.

3.1 Examples in real life

Renyi's entropy is most widely used in cryptographic applications, to measure the unpredictability or randomness of cryptographic keys. It also plays a crucial role in assessing the security of encryption algorithms.
In information retrieval systems, Renyi's entropy is used to optimize search algorithms and diversify results. It is also sometimes used in machine learning for clustering and classification tasks.

3.2 Properties

1. $H_\alpha(X)$ is non-negative:
$$H_\alpha(X) \geq 0$$
2. $H_\alpha(X)$ is additive by construction:
$$H_\alpha(A, B) = H_\alpha(A) + H_\alpha(B)$$
3. $H_\alpha(X)$ is concave for $\alpha < 1$ and for $\alpha > 1$, it is neither pure convex nor pure concave.
4. The entropies $H_\alpha(X)$ are monotonically decreasing with respect to the parameter α, i.e
$$\alpha < \alpha' \Rightarrow {H_\alpha}'(X) \geq H_\alpha(X)$$

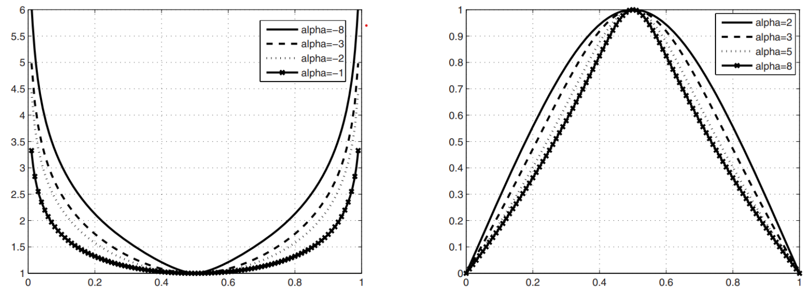

Figure 2: Renyi Entropy plots for negative and positive values of α

3.3 Varying the parameter α

For some specific values of α, $H_\alpha(X)$ has certain useful properties and thus has some particular names.

Shannon Entropy Consider the case when $\alpha \to 1$.

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

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

(using L'Hospitals rule)

$$\Rightarrow \lim_{\alpha \to 1} H_\alpha(X) = H_S(X)$$

where $H_S(X)$ represents the Shannon Entropy. Hence, we can see that Shannon Entropy is a special case of Renyi's Entropy.

Max-Entropy Consider the case when $\alpha = 0$.
Here, the Renyi Entropy equation reduces to:

$$H_0(X) = \frac{1}{1-0} \log(\sum_{i=1}^{N} (p_i)^0)$$

$$\Rightarrow H_0(X) = \log(\sum_{i=1}^{N} 1)$$

$$\Rightarrow H_0(X) = \log N$$

This is the maximum value the entropy can take. So, the shannon entropy of a probability distribution is never greater than the max-entropy.

The max-entropy is also called Hartley Entropy.

Collision Entropy In the case when $\alpha = 2$, we get the Collision Entropy of a probability distribution:

$$H_2(X) = \frac{1}{1-2} \log(\sum_{i=1}^{N} (p_i)^2$$

$$\Rightarrow H_2(X) = -\log(\sum_{i=1}^{N} p_i^2)$$

We see that this equation takes the form of the negative of the logarithm of the collision probability, that is, the probability that two independent random variables described by p will take the same value.

Min-Entropy The min-entropy describes the probability of the most likely outcome for a probability distribution.
It is never greater than the ordinary entropy of that distribution.

For a random variable X defined by probabilities $(p_1, p_2, ..., p_N)$, it's min-entropy is defined as:

$$H_min(X) = \log \frac{1}{p_max},$$

where

$$p_max = \max_i p_i.$$

When compared to Shannon Entropy, which is defined as

$$H(X) = \sum_i p_i \log(\frac{1}{p_i})$$

We can see that it can be read as the expectation value of $\log \frac{1}{p_i}$ over the distribution.

Subsequently, if the minimum value is taken instead of the expectation value, then we obtain the min-entropy definition.

3.4 Renyi's Relative Entropy

The Kullback–Leibler divergence (KL-divergence) or relative entropy of two probability distributions tells us how similar two distributions are. In other words, if relative entropy is zero then the two distributions are equivalent.

Similarly, we can define the Renyi divergence or α-divergence as

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \sum_{i=1}^{N} (P(i)^\alpha Q(i)^{1-\alpha})$$

Just like for entropy, when $\alpha \to 1$, the relative entropy also becomes equal to the KL-divergence.

$$\lim_{\alpha \to 1} D_\alpha(P\|Q) = \sum_{i=1}^{N} P(i) \log(\frac{P(i)}{Q(i)})$$

$$\Rightarrow \lim_{\alpha \to 1} D_\alpha(P||Q) = D_{KL}(P\|Q)$$

Also note that

$$D_\alpha(P\|Q) \geq 0$$

and the equality holds iff

$$P(X = x) = Q(X = x)$$

3.5 Renyi's Conditional Entropy

Similar to Shannon's entropy, it is only natural that we consider conditional entropy for Renyi's too.

Actually, several different definitions for Renyi's conditional entropy has been proposed, the most prominent of them being the one that was proposed by Renner and Wolf in 2005

$$R_\alpha^{RW}(X|Y) = \frac{1}{1 - \alpha} max_{y \in Y} \log \sum_{x \in X} P_{X|Y}(x|y)^\alpha$$

There are two more general definitions of conditional Renyi's entropy, which can be seen as analogues to conditional Shannon's entropies.

$$R_\alpha(X|Y) = \sum_{y \in Y} P_Y(y) R_\alpha(X|Y = y)$$

$$R_\alpha(X|Y) = R_\alpha(XY) - R_\alpha(Y)$$

3.6 Uniqueness Theorem

The uniqueness theorem tries to establish the fact that a function acting on a probability distribution, while following the Shannon-Khinchin axioms and exhibiting additivity, will take the form of Renyi's entropy.

As derived by Daroczy in his paper published 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

$$H(P) = G^{-1}\left(\frac{\sum_{i=1}^{N} f(p_i)G(-\log_2 p_i)}{\sum_{i=1}^{N} f(p_i)}\right) \tag{27}$$

where

$$\sum_{i=1}^{N} f(p_i) \leq 1,$$

and

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 $P = (p_1, ..., p_N)$ and $Q = (q_1, ..., q_M)$,

$$P * Q = (p_1q_1, ..., p_1q_M, ..., p_Nq_1, ..., p_Nq_M)$$

A.4. Provided that

$$\sum_{i=1}^{N} f(p_i) + \sum_{j=1}^{M} f(q_i) \leq 1,$$

we have

$$H(P \cup Q) = G^{-1}\left(\frac{\sum_{i=1}^{N} f(p_i)G(H(P)) + \sum_{j=1}^{M} f(q_j)G(H(Q))}{\sum_{i=1}^{N} f(p_i) + \sum_{j=1}^{M} f(q_j)}\right)$$

Now, we prove the uniqueness theorem for generalized entropies.

Theorem 2. *The function H(P) given in Equation (27) under the postulated A.1 to A.4 can have only one of the following two forms:*

$$H(P) = H_1^f(P) = -\frac{\sum_{i=1}^N f(p_i)\log_2 p_i}{\sum_{i=1}^N f(p_i)}$$

or,

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

Proof:
By taking n = 2 in Equation (2),

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

Using A.3, we have

$$H(p_1q, p_2q, ..., p_Nq) = H(p_1, p_2, ...p_N) + H(q)$$

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

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

Substituting $-\log p_i = x_i$ and $-\log_2 q = y$, then

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

where $\sum_{i=1}^N f(2^{-x_i}) \leq 1$. Now, we put $G(x_i+y) = G_y(x_i)$,

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

For this equation to be strictly monotonic and continuous, the following should be true for the relationship between $G_y(x)$ and $G(x)$:

$$G_y(x) = AG(x) + B$$

where $A \neq 0$ and B are arbitrary constants. 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

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

These two forms of the solution correspond to the two forms mentioned in the theorem when $c = (1 - \alpha)\log_e 2$.

Hence, the proof is complete.

3.7 Generalization to Quantum

In non-asymptotic or non-ergodic settings, where the law of large numbers fails to apply, entropy measures like the min-, the max-, or the collision entropy play an important role in analysing the system. Renyi's entropy nicely unifies these different and isolated measures into an α parameterized entropy measure.

A density matrix ρ in quantum physics, is a Hermitian, positive semidefinite matrix of trace one. Density matrices describe quantum physical systems in either mixed or pure states.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

If $p_{i\,i}$ are the eigenvalues of ρ, then the quantum Rényi entropy reduces to a Rényi entropy of a random variable X_ρ with probability distribution p_i

$$S_\alpha(\rho) = H_\alpha(X)$$

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.

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

The special cases of the Renyi's quantum entropy for α tends to zero and infinity are especially importnat in classifying quantum systems.

Minimum Value The Renyi entropy equals zero if and only if ρ is a pure state because a quantum system in a pure state will have $Tr(\rho) = 1$

Maximum Value The Renyi entropy is upper bounded by

$$S_\alpha(\rho) \leq log(d)$$

where d is the dimension of the complex Hilbert Space (vectors that represent physical states). Equality is achieved if and only if ρ is a maximally mixed state.

The quantum relative entropy of a state σ and positive semi-definite operator ρ is defined as

$$D(\sigma||\rho) = Tr[\sigma(\log\sigma - \log\rho)]$$

The Petz-Renyi relative entropy is one way to generalise the classical Renyi relative entropy to the quantum one.

$$D_\alpha(\sigma||\rho) = \frac{1}{\alpha - 1}\log(Tr[\sigma^\alpha \rho^{1-\alpha}])$$

As we can tell, when $\alpha \to 1$ the above quantity converges to the quantum relative entropy.

$$D(\sigma||\rho) = \lim_{\alpha\to 1} D_\alpha(\sigma||\rho)$$

Since most problems of quantum mechanics deals with quantifying the amount of information that is missing or unknown about a quantum state, entropy and relative entropy play a huge role in classifying the quantum entanglement (the phenomenon when quantum correlation between particles leads to non-classical correlations and behaviors) and the equilibrium of quantum systems.

4 Comparing Renyi's and Tsallis' Entropy

As mentioned before, both Renyi's and Tsallis' entropy are generalised forms of entropy that give different entropy distributions for different values of the paramter α and q respectively. Further, for lower values of the parameters α and q, both the entropies emphasize higher probabilities, while higher values focus more on the distribution's maximum probability.
There is also a relatively starightforward relation between Renyi's and Tsallis' entropies that can be derived from their equations:

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

or,

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

where R_α represents Renyi's Entropy and T_α represents Tsallis' Entropy. Also, from the very construction of Renyi's entropy, it is additive in nature and so it follows the relation $H_\alpha(AB) = H_\alpha(A) + H_\alpha(B)$. Thus, $H_\alpha(X)$ exhibits strong additivity.

On the other hand, Tsallis' entropy follows the relation of $H_q(A,B) = H_q(A) + H_q(B) + (1-q) * H_q(A) * H_q(B)$, and thus it exhibits weak additivity.

5 Applications

5.1 Using Tsallis linear estimator to detect Alzheimer's and Related diseases [5]

Estimating the information present in natural data, like electroencephalography (EEG) data, is particularly challenging because the connection between the physical data and the information it represents is not understood. This lack of understanding about how data translates to information is commonly referred to as the encoding problem. This is primarily due to, unlike other forms of encoding and decoding where symbols from a source set are encoded into a compressed binary representation and decoded based on codebooks (also known as symbolic encoding), EEG uses physical encoding where the waveform directly correlates with the brain's electrical activity. In EEG, the brain's neural signals are transformed into waveforms that represent various brain states or activities without an intermediary symbolic layer. This direct physical representation of brain activity makes interpreting and understanding EEG data more complex, as it requires a direct correlation between the observed waveforms and the underlying neural processes.
The authors find that all physical instantiation or action in the ECG occurs between points of discontinuity or critical points (maximum, minimum, or saddle point) in the data. They also notice that the variance within an EEG waveform is directly proportional to the probability of the occurrence of that discontinuous subsection.
So, based on the above findings, the authors propose a way to cut and bin the EEG waveform into separate information events. Then we use the frequency counts for waveforms and the variance information to develop a Tsallis entropy functional, allowing us to estimate without deducing the encoding. Using this approach, the accuracy for identifying individuals at risk of Alzheimer's and related diseases from brain EEG activity is 90%, which is one of the best in the literature.

5.1.1 Proofs and Mathematical formulation

The binning of EEG graph into subsections based on points of discontinuity stems from the fact that Neuronal encodings are somehow related to spikes. Although the exact nature of this encoding is unknown, it is generally agreed upon that these spikes are related to neural information.

Spikes are sharp upticks in the voltage of the neuron's axon, the action potential. These spikes appear as discontinuities in the voltage, the discontinuities primarily being peaks. From these observations and many others, this work hypothesizes that the physical encoding of information, that is, the values of yij, begin and end at local discontinuities and at unique critical points (local maximums, minimums, and saddle points) in the physical occurrences of the information.

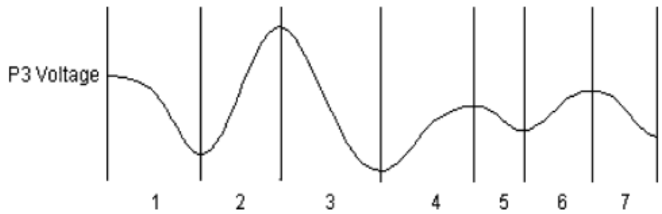

Fig. 2. An EEG waveform binned at critical points.

To prove the relation between variance and probabilities, we start with the power-frequency tradeoff, which says that the average power of an information element is inversely proportional to the probability of its occurrence. Based on this, Consider a large data set, which represents the occurrence of the specific events $\mathbf{y}_{ij}$. In this data set, each of the different types, j, is in proportion to their frequency of occurrence. Then the average power ($\frac{K}{p(\mathbf{y}_i)}$) is

$$\frac{K}{p\left(\mathbf{y}_i\right)} = \frac{\alpha}{N_{jk}} \sum_{j,k}^{N_{jk}} \left(y_{ijk} - \mu\right)^2 .$$

Here, α is a proportionality constant, and μ is the mean value of the entire voltage signal. Define $K' \equiv K/\alpha$, then $K' = \sigma^2/N$, where σ^2 is the average variance of the signal that carries the information and N is the number of different general events. We can write this summation as two separate terms, the variance within term and the variance between term:

$$\begin{aligned} \frac{K'}{p\left(\mathbf{y}_i\right)} &= \frac{\sigma^2}{Np\left(\mathbf{y}_i\right)} \\ &= \frac{1}{N_{jk}} \sum_{j,k}^{N_{jk}} \left(y_{ijk} - \mu\right)^2 \\ &= \frac{1}{N_{jk}} \sum_{j,k}^{N_{jk}} \left(y_{ijk} - \bar{y}_i\right)^2 + \left(\bar{y}_i - \mu\right)^2 . \end{aligned}$$

Here, $\bar{y}_i = 1/N_{jk} \sum_{j,k}^{N_{jk}} y_{ijk}$ is the local mean. Rewritten in simplified notation:

$$\frac{\sigma^2}{Np\left(\mathbf{y}_i\right)} = S_i^2 = \widehat{S}_i + \bar{S}_i^2 .$$

Using central limit theorem, we can show that $\bar{S}_i^2$ is distributed approximately as a chi-squared random variable with degree 1 and variance σ^2 and hence the approximate value is

$$\bar{S}^2\left(p\left(\mathbf{y}_i\right)\right) \simeq \sigma^2 \left(\frac{1}{Np\left(\mathbf{y}_i\right)} - p\left(\mathbf{y}_i\right)\right)$$

Substitution this into the simplified notation equation, we get

$$\begin{aligned}
\widehat{S}_i^2 &= \frac{\sigma^2}{Np(\mathbf{y}_i)} - \bar{S}_i^2 \\
&\simeq \frac{\sigma^2}{Np(\mathbf{y}_i)} - \sigma^2\left(\frac{1}{Np(\mathbf{y}_i)} - p(\mathbf{y}_i)\right) \\
&\simeq \sigma^2 p(\mathbf{y}_i)\,.
\end{aligned}$$

i.e. the probability of an event $p(y_i)$ is directly proportional to the average variance within $\widehat{S}_i^2$. Now, we can use this approach for a probability estimator $\widehat{p}(\mathbf{y}_i)$ and, along with the frequency $\widetilde{p}(\mathbf{y}_i)$, can estimate the Tsallis entropy.

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

This can also be simplified to

$$\widehat{H}_T(Y) = 1 - \frac{\sum_{i,j}^{M} \widehat{S}_{ij}^2}{S^2}$$

Where $\widehat{S}_{ij}^2$ is the variance within each interval and S^2 is an estimator for total variance σ^2.

5.1.2 Results

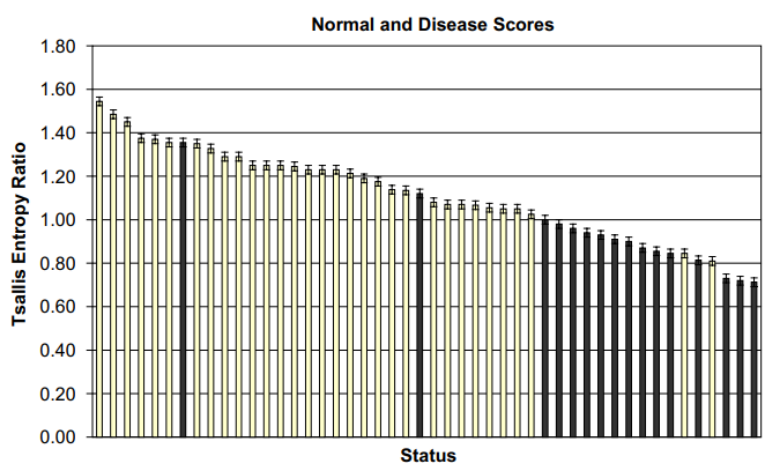

Fig. 5. Tsallis entropy ratio in ADRD and normal individuals.

Figure 3: The dark lines correspond to individuals with ADRD and light lines correspond to normal individuals

We see that Tsallis estimator is an efficient way to measure entropy in signals, Now we consider Tsallis entropy estimator for analyzing EEG data, particularly for the early detection of Alzheimer's disease and related disorders (ADRD). ADRD's primary symptom is memory loss, where sufferers may struggle to recognize faces seen recently, unlike those with healthy memory.

The hypothesis here is that this impaired recognition in ADRD patients is due to deficient information processing, suggesting that the necessary information for recall is often absent in individuals with ADRD.

Consequently, a proposed metric is the ratio of frontal brain information to posterior brain information. For healthy individuals, this ratio should be greater than 1, indicating higher frontal brain activity. In contrast, for those with ADRD, the ratio is expected to be less than 1, reflecting a decrease in frontal brain activity relative to the posterior part. The results are as above.

5.2 Image thresholding using Tsallis [7] and Renyi [8] entropy

Image segmentation is one of the most critical tasks in automated image analysis. It involves techniques to find objects that are not cleanly extractable by CV-based processes. Image thresholding is a broader class of such processes. One of the most efficient and popular techniques for image segmentation is entropy-based thresholding. This approach considers Shannon entropy and considers gray-level image histogram as the probability distribution. Tsallis and Renyi's entropy are also considered as a generalization of Shannon entropy and is explored in the field of image thresholding due to the desirable property of nonadditive information content in some image classes.

For image thresholding, first, the image is converted to an "image histogram," which is the frequency of each intensity level in a grayscale image. Then, the histogram is normalized to create a probability distribution. Entropy is a measure of randomness or unpredictability. For image thresholding, entropy is calculated for the foreground and background separately. The idea is to find a threshold that maximizes the sum of foreground entropy and background entropy. Here foreground entropy is calculated using the probabilities of the intensity levels below the threshold, and background entropy is calculated using the probabilities of the intensity levels above the threshold. Finally, the optimal threshold is chosen based on the maximum combined entropy achieved, and the segmented images are generated in a binary fashion by setting all values below the threshold as black (which typically represents the background) and all values equal and above the threshold to white (which is typically the foreground/ objects of interest).

Since it seems illogical to assume that just because two images might have similar grayscale distributions, the threshold values and the obtained foreground background images need not be similar. Hence, there has been a lot of exploration with regards to using some second-order statistics and entropy functions, including Tsallis and Renyi entropy.

5.2.1 Mathematical formulation for thresholding

Let $p_0, p_1, p_2, \ldots, p_{255}$ be the probability distribution of gray levels. From this distribution, two probability distributions, one for the object class A_1 and the other for the background class A_2, are derived. The probability distributions of the object and background classes, A_1 and A_2, are given by

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

and

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

where

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

This formulation is common for both approaches, but after this, there is a change in how entropy is calculated.

5.2.2 Tsallis Entropy

The a priori Tsallis entropy for each distribution is defined as

$$S_q^{\mathrm{A}_1}(t) = \frac{1 - \sum_{i=1}^{t} \left(\frac{p_i}{p^{\mathrm{A}_1}}\right)^q}{q-1}$$
$$S_q^{\mathrm{A}_2}(t) = \frac{1 - \sum_{i=t+1}^{k} \left(\frac{p_i}{p^{\mathrm{A}_2}}\right)^q}{q-1}$$

The Tsallis entropy $S_q(t)$ is parametrically dependent upon the threshold value t for the foreground and background. It is formulated as the sum of each entropy, allowing the pseudo-additive property for statistically independent systems. From $S_q(A+B) = S_q(A) + S_q(B) + (1-q) \cdot S_q(A) \cdot S_q(B)$, We get,

$$\begin{aligned} S_q(t) =& \frac{1 - \sum_{i=1}^{t} \left(p_{\mathrm{A}_1}\right)^q}{q-1} + \frac{1 - \sum_{i=t+1}^{k} \left(p_{\mathrm{A}_2}\right)^q}{q-1} \\ &+ (1-q) \cdot \frac{1 - \sum_{i=1}^{t} \left(p_{\mathrm{A}_1}\right)^q}{q-1} \cdot \frac{1 - \sum_{i=t+1}^{k} \left(p_{\mathrm{A}_2}\right)^q}{q-1} \end{aligned}$$

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.

$$t_{\mathrm{opt}} = \operatorname{argmax}\left[S_q^{\mathrm{A}_1}(t) + S_q^{\mathrm{A}_2}(t) + (1-q) \cdot S_q^{\mathrm{A}_1}(t) \cdot S_q^{\mathrm{A}_2}(t)\right]$$

5.2.3 Renyi Entropy

The Renyi's entropies associated with object and background distributions are given by

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

and

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

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

$$t^*(\alpha) = \operatorname{Arg} \max_{t \in G} \left\{H_{A_1}(t) + H_{A_2}(t)\right\}.$$

5.2.4 Results

In both of the cases, the parameters q and α are variable. Hence, apart from maximizing entropy and optimal thresholds, it is also crucial to vary over the secondary variables and pick the best performing. Some works mention the q parameter in Tsallis to the nonexistivity of physical systems, but might be possible to assume that q might have a physical meaning in the case of image segmentation and be determined based on constraints. The use of second-order entropy measures, i.e., both Tsallis and Renyi entropy, often perform better than Shannon entropy.

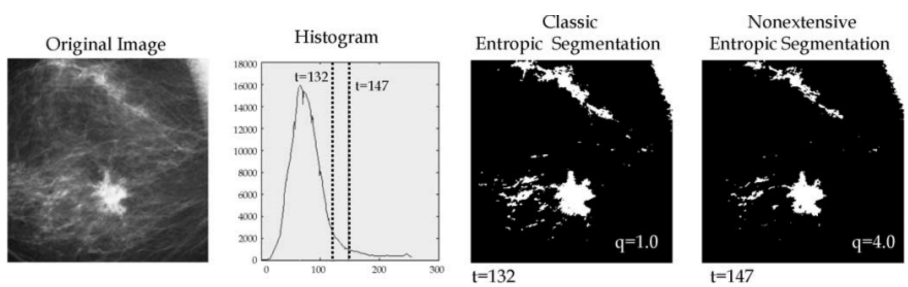

Figure 4: Example of entropic segmentation for mammography image with an inhomogeneous spatial noise. Two image segmentation results are presented for q = 1.0 (classic entropic segmentation) and q = 4:0.

5.3 Financial Modelling with Renyi's Entropy [6]

5.3.1 Introduction

In the financial world, knowledge about the future can be expressed in terms of possible events and their probabilities. In such situations, disagreements can be

understood as a mismatch between probability distributions. We can quantitatively measure these disagreements by the use of divergences, KL divergences, and Renyi's divergence.
Renyi's divergence was also more explored since it intuitively made sense to use a sensitive function like it since most financial data is precise to a large extent, which can be leveraged in Renyi's divergence.

5.3.2 Mathematical Formulation

We assume that m is the correct distribution (or the market's distribution regarding an event) and b is the investor's belief of a distribution for an experiment X; then Renyi's divergence between the two distributions is given by

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

Where the integration is done over all possible outcomes of the experiment X. Now if the investor knows about m, then based on his risk appetite R, we can compute the expected profit rate of the gamble between the distributions m and b from the formula.

$$Profit = \frac{1}{R} D_1(b||m) + \left(\frac{R-1}{R}\right) D_{1/R}(b||m)$$

The main reason why this is not possible in real life is primarily due to the availability of market odds for any event. Most and any type of prediction will already be a version of the investor's beliefs which makes it incredibly hard to predict events using this approach.

5.4 Other Applications of Tsallis and Renyi's Entropy

Apart from the aforementioned applications, Tsallis and Renyi's are used in multiple other fields. In fields of Topic Modeling Optimization, where both Tsallis and Renyi are used to determine the optimal number of topics in large textual collections while maintaining semantic similarities across topics [9]. They are also used in Machine learning as a substitute for Shannon entropy, which is more prevalent, especially in generations of decision trees and for cross-entropy functions [11] [13]. There is also work regarding network complexity measures, a combination of local and global-based metrics derived from Tsallis entropy [15]. In conclusion, Since both Tsallis and Renyi's entropies are generalizations of Shannon, they are often used in specialized systems where the rigidness of Shannon entropy affects accurate modeling and hence are quite prevalent in niche topics of interest.

6 Bibliography

References

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

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

[3] 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.

[4] 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.

[5] Sneddon, R. (2007). The Tsallis entropy of natural information. Physica A: Statistical Mechanics and Its Applications, 386(1), 101–118.

[6] Soklakov AN. Economics of Disagreement-Financial Intuition for the Rényi Divergence. Entropy (Basel). 2020 Aug 3;22(8):860. doi: 10.3390/e22080860. PMID: 33286632; PMCID: PMC7517462.

[7] Portes de Albuquerque, M., Esquef, I. A., Gesualdi Mello, A. R., & Portes de Albuquerque, M. (2004). Image thresholding using Tsallis entropy. Pattern Recognition Letters, 25(9), 1059–1065. doi:10.1016/j.patrec.2004.03.003

[8] Sahoo, P., Wilkins, C., & Yeager, J. (1997). Threshold selection using Renyi's entropy. Pattern Recognition, 30(1), 71–84. doi:10.1016/s0031-3203(96)00065-9

[9] Sergei, K. (2018). Application of Rényi and Tsallis Entropies to Topic Modeling Optimization. ArXiv. /abs/1802.10526

[10] Alberto Enciso, Piergiulio Tempesta (2017). Uniqueness and characterization theorems for generalised entropies, arXiv:1702.01336

[11] Maszczyk, T., Duch, W. (2008). Comparison of Shannon, Renyi and Tsallis Entropy Used in Decision Trees. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds) Artificial Intelligence and Soft Computing – ICAISC 2008. ICAISC 2008. Lecture Notes in Computer Science(), vol 5097. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69731-2_62

[12] Tim van Erven, Peter Harremoës (2014), Rényi Divergence and Kullback-Leibler Divergence, arXiv:1206.2459

[13] Ramezani, Z., & Pourdarvish, A. (2020). Transfer learning using Tsallis entropy: An application to Gravity Spy. Physica A: Statistical Mechanics and Its Applications, 125273. doi:10.1016/j.physa.2020.125273

[14] José M.Amigo, Sámuel G.Balogh, Sergio Hernández (2018), A Brief Review of Generalized Entropies, https://doi.org/10.3390/e20110813

[15] Wen, T., & Jiang, W. (2019). Measuring the complexity of complex network by Tsallis entropy. Physica A: Statistical Mechanics and Its Applications, 526, 121054. doi:10.1016/j.physa.2019.121054

7 Contributions

1. Rohan: Introduction to Tsallis entropy, properties, relation with shannon entropy and uniqueness theorem.
2. Paridhi: Introduction, Renyi's entropy and properties, Uniqueness Theorem for Renyi's Entropy, Chain Rule For Renyi's entropy
3. Sreeja: Uniqueness Theorem for Renyi's Entropy, Generalisation of Renyi's Entropy to Quantum
4. Harshavardhan: Applications