Axiomatic Definition of Entropy

Course: *CS258 Information Theory*

Author: 丁健宇 *519030910010*

Author: 张若涵 *519030910029*

Author: 易文龙 *519030910068*

Professor: *Fan Cheng*

Date: *October 26, 2023*

Contents

0. Preface

We all know that to a discrete random variable X defined in $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$, its information entropy is defined as

$$H(X) = H(\mathcal{P}) = -\sum_{i=1}^{n} p_i \log p \tag{0.1}$$

But what if another function H' on $\mathcal{P}$ satisfies a series of properties, which are consider to be the basic feature of information entropy. Can we call H' to be the entropy function?

1. Probability space and General measure space

A measure space is a basic object of measure theory, a branch of mathematics that studies generalized notions of volumes. It contains an underlying set, the subsets of this set that feasible for measuring (the σ-algebra) and the method that is used for measuring (the measure). For example, we can define $\mathcal{P}$ as a measure on a measure space $(\Omega, \mathcal{F})$
Actually, one important example of a measure space is a probability space. In probability theory, a probability space or a probability triple $(\Omega, \mathcal{F}, P)$ is a mathematical construct that provides a formal model of a random process or "experiment". Conclusively, we will get a probability space when we put the concept of probability into the measure space. As is known to us, the definition of the σ-algebra should satisfy CUT, which refers to Complement, Union and Total Set. Similarly, we should obey the rule of CAT when defining the possibility, namely Countable Additivity and Total Set. Specifically, the denotation of the total set is that the possibility of the whole events is 1; and Countable Additivity demonstrates that for any subsets of the whole events, once they are not intersected, which means:

$$\forall i \neq j, \mathcal{A}_i \cap \mathcal{A}_j = \varnothing$$

We can definitely draw the conclusion that:

$$\mathcal{P}(\cup_{k=1}^{\infty} \mathcal{A}_k) = \sum_{k=1}^{\infty} \mathcal{P}_k$$

2. Rényi' s Axioms in probability space

In this section, we focus on the existence of forementioned H' in probability space, *i.e.* whether H is unique when it satisfies the following Rényi's Axioms. Note that the axiom of *Smoothness* is removed because the density function f in probability space is on a scale of $[0, 1]$.

Given a finite set $X = \{x_1, x_2, ..., x_n\}$, two signed probability measure $P = \{p_1, p_2, ..., p_m\}$ and $Q = \{q_1, q_2, ..., q_n\}$, the following axioms on imposed on entropy H.

Axiom 1. *(Real-Valuedness) $H(\mathcal{P})$ is a non-constant real-valued function on $\mathcal{P}$.*

Axiom 2. *(Symmetry)* $H(\mathcal{P})$ *is a symmetric function of the elements of* $\mathcal{P}$.

Axiom 3. *(Continuity)* $H(\mathcal{P})$ *is continuous of each of elements of of* $\mathcal{P}$.

Axiom 4. *(Calibration)* $H((1/2)) = 1$.

Axiom 5. *(Additivity)* $H(\mathcal{P} * \mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{Q})$

where the operation of $\mathcal{P} * \mathcal{Q}$ is defined as

$$\mathcal{P} * \mathcal{Q} := \cup_{p_i \in \mathcal{P}, q_j \in \mathcal{Q}} \{p_i q_j\} \tag{2.1}$$

Axiom 6. *(Mean-Value Property)* $\exists g : \mathbb{R} \to \mathbb{R}$, *that*

$$H(\mathcal{P} \cup \mathcal{Q}) = g^{-1}\left[\frac{w(\mathcal{P})g(H(\mathcal{P})) + w(\mathcal{Q})g(H(\mathcal{Q}))}{w(\mathcal{P} \cup \mathcal{Q})}\right] \tag{2.2}$$

where the quantity $w(Q) =\mid \sum_i q_i \mid$ is called as the weight of Q. We require $w(Q) \neq 0$ but we do not require $w(Q) = 1$ (except when Q is a signed probability measure).

3. Deduction of Rényi entropy in probability space

In this section, we suppose that the form of entropy H is not unique and we are going to prove the following theorem to validate our supposition.

Theorem 1. *Axioms 1-6 hold if and only if*

$$H(\mathcal{P}) := H_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha\right) \tag{3.1}$$

where α *is a free parameter that* $\alpha > 0$.

In order to prove such a concrete theorem, four lemmas should be introduced first.

Lemma 1. *Under Axioms 5* $\forall q \neq 0, H((q)) = -c_q log|q|$, *where* $c_q \in \mathbb{R}$ *satisfies that* $\forall q_1/q_2 \in \mathbb{Q}, c_{q_1} = c_{q_2}$ *and* H *can be represented according to* ***Axiom of Choice***.

Proof. Let $h(q) = H((q))$.
Axiom 5 implies that

$$h(pq) = h(p) + h(q) \tag{3.2}$$

whenever $p, q \neq 0$.
Hence

$$\forall n \in \mathbb{Z}^*, h(q^n) = H(n(q)) = nh(q) \tag{3.3}$$

And therefore

$$\forall r \in \mathbb{R}^*, h(q^r) = rh(q) \tag{3.4}$$

□

Lemma 2. *Under Axioms 1, 3, 4, and Lemma 1,* $\forall q \neq 0, H((q)) = -log_2|q|$.

Proof. Based on Axiom 3, a non-zero rational sequence $\{r_i\}$ can be chosen in a complete field $\mathbb{F}$ that converges to $x \in \mathbb{F}^*$, i.e. $h(r_n) \to h(x)$. Thus it comes that

$$\forall q \neq 0, H((q)) = clog_2|q|, where\ c\ \in \mathbb{F}.$$

Axioms 1 and 4 together imply that

$$\forall q \neq 0, H((q)) = -log_2|q| \tag{3.5}$$

□

Lemma 3. *(Hardy, Littlewood, Pólya [1952, Theorem 83])*
Mappings which generate the same means are linearly related.
i.e. Assuming that $\sum c_n = 1, c_n > 0$, $\phi(x)$ *is a strictly monotone and continuous function, weighted mean* $\mathfrak{R}_\phi(a) := \phi^{-1}[\sum_i c_i\phi(a_i)]$.
Therefore $\mathfrak{R}_\chi(a) \equiv \mathfrak{R}_\psi(a)$ *iff* $\exists\ \alpha$ *and* β *s.t.*$\chi = \alpha\psi + \beta$.

Proof. First considering " $\Rightarrow$ ". $\forall t \in [H, K]$,

$$\begin{aligned} x &:= \psi^{-1}\left[\frac{K-t}{K-H}\psi(H) + \frac{t-H}{K-H}\psi(K)\right] \\ &= \chi^{-1}\left[\frac{K-t}{K-H}\chi(H) + \frac{t-H}{K-H}\chi(K)\right] \end{aligned} \tag{3.6}$$

Then if t traverses (H, K), x can be every value in (H, K).
Consequently,

$$\begin{aligned} \chi(x) &= \frac{K-t}{K-H}\chi(H) + \frac{t-H}{K-H}\chi(K) \\ &= \frac{\psi(K)-\psi(x)}{\psi(K)-\psi(H)}\chi(H) + \frac{\psi(x)-\psi(H)}{\psi(K)-\psi(H)}\chi(K) \\ &= \alpha\psi(x) + \beta \end{aligned} \tag{3.7}$$

" $\Rightarrow$ " has been proved and " $\Leftarrow$ " obviously holds true. □

Lemma 4. *Under Axioms 5, 6 and Lemma 3,* g *is a linear function* $g(x) = -ax + b$ *or exponential function* $g(x) = a \cdot 2^{(1-2k)x}$, *where* $k \in \mathbb{N}^+$.

Proof. Axiom 6 implies that

$$\begin{aligned} H(\mathcal{P}) &= H((p_1) \cup \cdots \cup (p_n) \cup \cdots) \\ &= g^{-1}\left[\frac{\sum_i w((p_i))g(H((p_i)))}{w((p_1) \cup \cdots \cup (p_n) \cup \cdots)}\right] \\ &= g^{-1}\left[\frac{\sum_i |p_i| g(-\log_2|p_i|)}{\sum_i |p_i|}\right] \end{aligned} \tag{3.8}$$

Define $f : \mathbb{R}_+ \to \mathbb{R}, t \mapsto g(-\log_2 t)$, it is easy to figure out that g is monotone and continuous.
According to Axiom 5, with $H(\mathcal{P} * (q))$ taken into account,

$$f^{-1}\left[\frac{\sum_i |p_i| f(|p_i q|)}{|\sum_i p_i|}\right] = |q| f^{-1}\left[\frac{\sum_i |p_i| f(|p_i|)}{|\sum_i p_i|}\right] \tag{3.9}$$

Let $h_q(t) = f(|q|t)$, and then

$$h_q^{-1}\left[\frac{\sum_i |p_i| h_q(|p_i|)}{|\sum_i p_i|}\right] = f^{-1}\left[\frac{\sum_i |p_i| f(|p_i|)}{|\sum_i p_i|}\right] \tag{3.10}$$

The conclusion can be drawn that h_q and f generate the same means.
Applying Lemma 3, f and h_q is linearly related, *i.e.* $h_q(t) = f(|q|t) = s(|q|)f(t) + r(|q|)$.

Here $s(|q|)$ is a constant associated with $r(|q|)$ and $|q|$, $s(|q|) > 0$ and then $f(|q|) = r(|q|)$.

Replacing q with y,

$$\begin{cases} f(ty) = s(y)f(t) + f(y) \\ f(ty) = s(t)f(y) + f(t) \end{cases} \tag{3.11}$$

and therefore $\dfrac{s(t)-1}{f(t)} = \dfrac{s(y)-1}{f(y)}$.
This implies that $\exists$ a constant c *s.t.* $s(t) = 1 + cf(t)$.
Substituting,

$$f(ty) = cf(t)f(y) + f(t) + f(y) \tag{3.12}$$

The functional equation has two solutions:

- $c = 0$, the solution is $f = C \log t$.
- $c \neq 0$, the original equation can be simplified into $[cf(ty) + 1] = [cf(t) + 1][cf(y) + 1]$.
  The solution is that $f = \dfrac{x^\alpha - 1}{c}$, where α can be any constant.

To summarize, $g(x) = -ax + b$ or $a \cdot 2^{(1-\alpha)x} + b$.
Recalling Theorem 1, now we can prove it with the aforementioned four lemmas.
Based on $H((q, 1-q)) = H((q) \cup (1-q))$, two solutions corresponding to different g can be derived.

- If $g(x) = -ax + b, -aH((q, 1-q)) + b = a[q \log_2 q + (1-q)\log_2(1-q)] + b$.
  In the other word, $H((q, 1-q)) = -q \log q - (1-q)\log(1-q)$ and obviously $H \in C^\infty(Q)$.
  This shows that

$$H(\mathcal{P}) = -\frac{\sum_i |p_i| \log_2 |p_i|}{|\sum_i p_i|} \tag{3.13}$$

- If $g(x) = a \cdot 2^{(1-\alpha)x} + b$,

$$a \cdot 2^{(1-\alpha)H(\mathcal{P})} + b = \frac{a\sum_i |p_i|^\alpha + b\sum_i |p_i|}{|\sum_i p_i|} \tag{3.14}$$

Let $\mathcal{P} = (q, 1-q)$. Since $\dfrac{|q| + |1-q|}{|q+1-q|}$ is not differentiable at $q = 0$, hence $b = 0$.
In consequence

$$H((q, 1-q)) = \frac{\log_2[|q|^\alpha + |1-q|^\alpha]}{1-\alpha} \tag{3.15}$$

Apparently $\alpha < 0$ results in an unbounded $H((0,1))$, thus aborted.
Simultaneously, Axiom 1 requires $\alpha \neq 0$, so $\alpha > 0$.
And thus

$$H_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} p_i^\alpha \right), \ \alpha > 0 \tag{3.16}$$

Note that equation (3.13) can also be generalized into equation (3.16) as $\alpha \to 1$.
Consequently

$$H_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} p_i^\alpha \right), \ \alpha > 0 \tag{3.17}$$

□

4. Rényi' s Axioms in general measure space

4.1. Definition in general measure space.

In general measure space, the axiom of *Smoothness* should be added when figuring out $H(\mathcal{P})$ in measure space.

Axiom 7. *$H((q, 1-q))$ is smooth at $q = 0$.*

4.2. Deduction of Rényi' s Axioms in general measure space.

Theorem 2. *Axioms 1-7 hold if and only if*

$$H(\mathcal{P}) = H_{2k}(\mathcal{P}) = -\frac{1}{2k-1} \log_2 \left[\frac{\sum_i |p_i|^{2k}}{|\sum_i p_i|} \right] \tag{4.1}$$

where $k = 1, 2, ...$ is a free parameter.

Proof. Now use Axiom 7.
Assuming that α in equation (3.16) is not an integer.
Setting $H(\mathcal{P}) = H((q, 1-q))$.
Given $q \in \mathbb{R}_+$,

$$\frac{\mathrm{d}H((q, 1-q))}{\mathrm{d}q} = \frac{\alpha}{(\alpha-1)\ln 2} \cdot \frac{q^{\alpha-1} + (1-q)^{\alpha-1}}{q^\alpha + (1-q)^\alpha} \tag{4.2}$$

It can be easily derived that

$$\frac{\mathrm{d}^n H((q,1-q))}{\mathrm{d}q^n} = C_n \cdot \frac{p_1^{(n)}(q) + p_2^{(n)}(q)[q^{\alpha-n} + (1-q)^{\alpha-n}]}{p_2^{(n)}(q)} \tag{4.3}$$

where $p_k^{(n)}(q)$ is a polynomial constituted by $q^{\alpha-m} + (1-q)^{\alpha-m}, \quad \mathbb{Z} \ni m < n$.

In a result, n should be a constant larger than α, and then $\frac{\mathrm{d}^n H((q,1-q))}{\mathrm{d}q^n} = \infty$, which leads to a contradiction.

In summary, α must be an integer.

Meanwhile, α should be an even integer, otherwise $\frac{\mathrm{d}^n H((q,1-q))}{\mathrm{d}q^n}$ is not equal at $q = 0^-$ and $q = 0^+$. To sum up,

$$H(\mathcal{P}) = H_{2k}(\mathcal{P}) = -\frac{1}{2k-1}\log_2\left[\frac{\sum_i |p_i|^{2k}}{|\sum_i p_i|}\right] \tag{4.4}$$

where $k \in \mathbb{N}^+$.

□

5. Different types of Rényi entropy

WLOG. We can look in to the discrete Rényi entropy of different order α given the definition as:

$$H_\alpha(\mathcal{P}) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n} p_i^\alpha\right), \ \alpha > 0 \tag{5.1}$$

We can rewrite it into the form of

$$H_\alpha(\mathcal{P}) = \frac{\alpha}{1-\alpha}\log||\mathcal{P}||_\alpha \tag{5.2}$$

where $||X||_p$ means p-norm of X.

5.1. Shannon entropy.
We can take the limitation of α as $\alpha \to 1$.

$$\begin{aligned} H_1(\mathcal{P}) &= \lim_{\alpha\to 1} H_\alpha(\mathcal{P}) \\ &= \lim_{\alpha\to 1} \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha\right) \\ &= \lim_{\alpha\to 1} \frac{1}{1-\alpha} \log\left(1 + \sum_{i=1}^{n} p_i(p_i^{\alpha-1} - 1)\right) \\ &\to \lim_{\alpha\to 1} \frac{1}{(1-\alpha)\ln 2} \cdot \sum_{i=1}^{n} p_i(p_i^{\alpha-1} - 1) \\ &= \lim_{\alpha\to 1} \frac{1}{\ln 2} \sum_{i=1}^{n} \left(p_i \cdot \frac{p_i^{\alpha-1} - 1}{1-\alpha}\right) \\ &\to \frac{1}{\ln 2} \sum_{i=1}^{n} (-p_i \log p_i) \\ &= -\sum_{i} (p_i \log_2 p_i) \end{aligned} \tag{5.3}$$

So the result reduce to Shannon entropy, and the relative properties we have been studied for a long time.

5.2. Hartley or Max-entropy.
If $\alpha = 0$,

$$H_0(\mathcal{P}) = \log n = \log |\mathcal{P}| \tag{5.4}$$

That is to say no matter how the variable distributed, the entropy depends only on the scale of $\mathcal{P}$. The entropy is called max-entropy because it's the maximum of Shannon code.

5.3. Collision entropy.
If $\alpha = 2$,

$$H_2(X) = -\log \sum_i p_i^2 = -\log P(X = Y) \tag{5.5}$$

where X and Y are independent and identically distributed.

The collision entropy describe the chance of a random variable to colliding with itself, which is also called "Index of coincidence". In cryptography, coincidence counting is the technique (invented by William F. Friedman[4]) of putting two texts side-by-side and counting the number of times that identical letters appear in the same position in both texts. This count, either as a ratio of the total or normalized by dividing by the expected count for a random source model, is known as the index of coincidence, or IC for short.

5.4. Min-entropy.
If $\alpha = \infty$,

$$H_\infty(X) = \min_i(-\log p_i) = -\log(\max_i p_i) \tag{5.6}$$

Contrary to max-entropy, the min-entropy describe the minimum entropy of a random variable X, which is not affected by the distribution or $|\mathcal{X}|$, but depend on the maximum possibility of $\mathcal{P}$.

6. Information Theory in the context of Rényi entropy

Based on the determined Rényi entropy, key concepts can be well defined, which establishes the Information Theory in the context of Rényi entropy.

6.1. Discrete Rényi entropy.
Let X be a discrete random variable with alphabet $\mathcal{X}$ and $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$. The α order of Rényi entropy of $\mathcal{X}$ is defined by

$$H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_i p_i^\alpha\right) = \frac{\alpha}{1-\alpha}\log||\mathcal{P}||_\alpha \tag{6.1}$$

The discrete Rényi entropy is also written as $H_\alpha(\mathcal{P})$.

6.1.1. Joint Rényi entropy.
Let X and Y be two discrete random variables, the joint Rényi entropy of them should be

$$H_\alpha(X,Y) = \frac{1}{1-\alpha}\log\left(\sum_{x,y} p^\alpha(x,y)\right) \tag{6.2}$$

6.1.2. Conditional Rényi entropy.
Let X and Y be two discrete random variables, the conditional Rényi entropy of them should be

$$\begin{aligned} H_\alpha(Y|X) &= \sum_x p(x) H_\alpha(Y|X = x_i) \\ &= \frac{1}{1-\alpha}\left(\sum_x p(x)\log\left(\sum_y p^\alpha(y|X=x)\right)\right) \end{aligned} \tag{6.3}$$

We may find that under most cases where $\alpha \neq 1$

$$\begin{aligned} H_\alpha(X,Y) - H_\alpha(X) &= \frac{1}{1-\alpha}\log\left(\sum_{x,y} p^\alpha(x,y)\right) - \frac{1}{1-\alpha}\log\left(\sum_i p_i^\alpha\right) \\ &= \frac{1}{1-\alpha}\left(\log\sum_{x,y}\frac{p^\alpha(x,y)}{\sum_x p^\alpha(x)}\right) \\ &= \frac{1}{1-\alpha}\left(\log\sum_x\left(\frac{p(x)}{||\mathcal{P}||_\alpha}\right)^\alpha\sum_y p^\alpha(y|X=x)\right) \\ &\neq H_\alpha(Y) \end{aligned}$$

6.1.3. Relative Rényi entropy.

The relative entropy of Rényi entropy is different from Shannon entropy(KL-divergence).

So the Rényi divergence between $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ and $\mathcal{Q} = \{q_1, q_2, \ldots, q_n\}$ over the alphabet $\mathcal{X}$ is defined as

$$D(P\|Q) = \frac{1}{1-\alpha} \log \left(\sum_{i,j} p_{i,j} (p_{i,j}/q_{i,j})^{\alpha-1} \right) \tag{6.4}$$

6.1.4. Mutual information.

$$\begin{aligned} I(X;Y) &= D(P(X,Y)||P(X)P(Y)) \\ &= \frac{1}{1-\alpha} \log \left(\sum_{x,y} p(x,y)(p(x,y)/(p(x)p(y))^{\alpha-1} \right) \end{aligned} \tag{6.5}$$

6.2. Differential Rényi entropy.

The differential entropy $h(X)$ of a continuous random variable X with density $f(x)$ is defined as

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \int_S f^\alpha(x) \mathrm{d}x = \frac{\alpha}{1-\alpha} \log \|f\|_\alpha \tag{6.6}$$

where S is the support set of the random variable. The differential Rényi entropy is also written as $h_\alpha(f)$.

6.2.1. Joint Rényi entropy.

Let X and Y be two discrete random variables, the joint Rényi entropy of them should be

$$h_\alpha(X,Y) = \frac{1}{1-\alpha} \log \left(\int_x \int_y f^\alpha(x,y) \right) \tag{6.7}$$

6.2.2. Conditional Rényi entropy.

Let X and Y be two random variables, the conditional Rényi entropy of them should be

$$\begin{aligned} h_\alpha(Y|X) &= \int_x f(x) h_\alpha(Y|X = x_i) \\ &= \frac{1}{1-\alpha} \left(\int_x f(x) \log \left(\int_y g^\alpha(y|X=x) \right) \right) \end{aligned} \tag{6.8}$$

6.2.3. Relative Rényi entropy.

The relative entropy of Rényi entropy is different from Shannon entropy(KL-divergence). So the Rényi divergence is defined as

$$D(f\|g) = \frac{1}{1-\alpha} \log_2 \int_{X,Y} f(x,y)(f(x,y)/g(x,y))^{\alpha-1} \mathrm{d}x\mathrm{d}y \tag{6.9}$$

6.2.4. Mutual information.

$$I(X;Y) = D(P(X,Y)\|P(X)P(Y)) \\ = \frac{1}{1-\alpha}\log\left(\int_{X,Y} p(x,y)(p(x,y)/(f(x)g(y))^{\alpha-1}\right) \tag{6.10}$$

7. The application of Rényi entropy

7.1. Quantum entropies.

For a density matrix $\rho \in D(\mathcal{H})$, the quantum Rényi entropy is defined as follows:

$$S_\alpha(\rho) = \frac{1}{1-\alpha}\log \mathrm{Tr}\left(\rho^\alpha\right),\ \alpha \in (0,1)\cup(1,\infty) \tag{7.1}$$

This is a quantum version of a classical Rényi entropy. 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 $\{p_i\}$

$$S_\alpha(\rho) = H_\alpha(X_\rho) = \frac{1}{1-\alpha}\log\left(\sum_i p_i^\alpha\right) \tag{7.2}$$

Rényi entropy is widely used in information theory, for example, in restricting error probabilities in classification problems[2], entanglement-assisted local operations and classical communications conversion[3], strong converse problem in quantum hypothesis testing[6], and strong converse problem for the classical capacity of a quantum channel[7].

7.2. Theoretical computer science.

In Theoretical computer science, Rényi entropy is used in the randomness extractor, often simply called an "extractor", is a function, which being applied to output from a weakly random entropy source, together with a short, uniformly random seed, generates a highly random output that appears independent from the source and uniformly distributed[9]. One important example of extractor is Von Neumann extractor, which is based on Rényi entropy when $\alpha \to \infty$

From the input stream, his extractor took bits, two at a time (first and second, then third and fourth, and so on). If the two bits matched, no output was generated. If the bits differed, the value of the first bit was output. The Von Neumann extractor can be shown to produce a uniform output even if the distribution of input bits is not uniform so long as each bit has the same probability of being one and there is no correlation between successive bits.

Thus, it takes as input a Bernoulli sequence with p not necessarily equal to $1/2$, and outputs a Bernoulli sequence with $p = 1/2$. More generally, it applies to any exchangeable sequence—it only relies on the fact that for any pair, 01 and 10 are equally likely: for independent trials, these have probabilities $p\cdot(1-p) = (1-p)\cdot p$, while for an exchangeable sequence the probability may be more complicated, but both are equally likely.

7.3. Ecology and Statistics.
In Ecology and Statistics, Rényi entropy can be used to describe diversity index (also called phylogenetic index) is a quantitative measure that reflects how many different types (such as species) there are in a dataset (a community), and that can simultaneously take into account the phylogenetic relations among the individuals distributed among those types, such as richness, divergence or evenness. These indices are statistical representations of biodiversity in different aspects (richness, evenness, and dominance).

The idea is that the more different letters there are, and the more equal their proportional abundances in the string of interest, the more difficult it is to correctly predict which letter will be the next one in the string. The Rényi entropy quantifies the uncertainty (entropy or degree of surprise) associated with this prediction.

7.4. Other field.
In the Heisenberg XY spin chain model, the Rényi entropy as a function of α can be explicitly calculated because it is a self-justifiable function of a particular subgroup of modular groups.

We can say that all the applications of Shannon's entropy can be generalized to the applications of Rényi's entropy, for Riley's entropy is a generalization of Shannon's entropy.

References

[1] S. S Chehade and A. Vershynina. Quantum entropies. *Scholarpedia*, 14(2):53131, 2019. revision #190401.

[2] I. Csiszar. Generalized cutoff rates and renyi's information measures. *IEEE Transactions on Information Theory*, 41(1):26–34, 1995.

[3] J. Cui, M. Gu, L. C. Kwek, M. F. Santos, H. Fan, and V. Vedral. Quantum phases with differing computational power. *Nature Communications*, 3(3):812, 2012.

[4] W. F. Friedman. The index of coincidence and its applications in cryptanalysis. 1922.

[5] Peter Jizba and Toshihico Arimitsu. Observability of renyi's entropy. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 69:026128, 03 2004.

[6] F. D. Nobre, M. A. Rego-Monteiro, and C. Tsallis. Nonlinear relativistic and quantum equations with a common type of solution. *Phys. Rev. Lett.*, 106:140601, Apr 2011.

[7] Masanori Ohya and Igor V. Volovich. On quantum capacity and its bound. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 06(02):301–310, 2003.

[8] A. Rényi. On the foundations of information theory. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 33(1):1–14, 1965.

[9] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, page 32, USA, 2000. IEEE Computer Society.