

Entropy Measures *vs.* Kolmogorov Complexity

Andreia Teixeira

FCUP and Instituto de Telecomunicações

Armando Matos

FCUP and LIACC

André Souto

FCUP and Instituto de Telecomunicações

Luís Antunes

FCUP and Instituto de Telecomunicações

Abstract

Kolmogorov complexity and Shannon entropy are conceptually different measures. However, for any recursive probability distribution, the expected value of Kolmogorov complexity equals its Shannon entropy, up to a constant. We study if a similar relationship holds for Rényi and Tsallis entropies of order α , showing that it only holds for $\alpha = 1$. Regarding a time-bounded analogue relationship, we show that, for some distributions we have a similar result. We prove that, for universal time-bounded distribution $\mathbf{m}^t(x)$, Tsallis and Rényi entropies converge if and only if α is greater than 1. We also establish the uniform continuity of these entropies.

1 Introduction

The Kolmogorov complexity $K(x)$ measures the amount of information contained in an individual object (usually a string) x , by the size of the smallest program that generates it. It naturally characterizes a probability distribution over Σ^* (the set of all finite binary strings), assigning a probability of $c \cdot 2^{-K(x)}$ for any string x , where c is a constant. This probability distribution is called universal probability distribution and it is denoted by \mathbf{m} .

The Shannon entropy $H(X)$ of a random variable X is a measure of its average uncertainty. It is the smallest number of bits required, on average, to describe x , the output of the random variable X .

Kolmogorov complexity and Shannon entropy are conceptually different, as the former is based on the length of programs and the later in probability distributions. However, for any recursive probability distribution (*i.e.*, distributions that are computable by a Turing machine), the expected value of the Kolmogorov complexity equals the Shannon entropy, up to a constant term depending only on the distribution (see (1)).

Several information measures or entropies have been introduced since Shannon's seminal paper (2). We are interested in two different generalizations of Shannon entropy:

- Rényi entropy (3), an additive measure based on a specific form of mean of the elementary information gain: instead of using the arithmetic mean, Rényi used the Kolmogorov-Nagano mean associated with the function $f(x) = c_1 b^{(1-\alpha)x} + c_2$ where c_1 and c_2 are constants, b is a real greater than 1 and α is a non-negative parameter;
- Tsallis entropy (4), a non additive measure, often called a non-extensive measure, in which the probabilities are scaled by a positive power α , that may either reinforce the large (if $\alpha > 1$) or the small (if $\alpha < 1$) probabilities.

Let $R_\alpha(P)$ and $T_\alpha(P)$ denote, respectively, the Rényi and Tsallis entropies associated with the probability distribution P . Both are continuous functions of the parameter α and both are (quite different) generalizations of the Shannon entropy, in the sense that $R_1(P) = T_1(P) = H(P)$ (see (5)). It is well known (see (6)) that for any recursive probability distribution P over Σ^* , the average value of $K(x)$ and the Shannon entropy $H(P)$ are close, in the sense that

$$0 \leq \sum_x P(x)K(x) - H(P) \leq K(P) \quad (1)$$

where $K(P)$ is the length of the shortest program that describes the distribution P . We study if this property also holds for Rényi and Tsallis entropies. The answer is no. If we replace H by R or by T , the inequalities (1) are no longer true (unless $\alpha = 1$). We also analyze the validity of the relationship (1), replacing Kolmogorov complexity by its time-bounded version, proving that it holds for distributions such that the cumulative probability distribution is computable in an allotted time. So, for these distributions, Shannon entropy equals the expected value of the time-bounded Kolmogorov complexity. We also study the convergence of Tsallis and Rényi entropies of the universal time-bounded distribution $\mathbf{m}^t(x) = c \cdot 2^{-K^t(x)}$, proving that both entropies converge if and only if $\alpha > 1$. Finally, we prove the uniform continuity of Tsallis, Rényi and Shannon entropies.

The rest of this paper is organized as follows. In the next section, we present the notions and results that will be used. In Section 3, we study if the inequalities (1) can be generalized for Rényi and Tsallis entropies and we also establish a similar relationship for the time-bounded Kolmogorov complexity. In Section 4, we analyze the entropies of the universal time-bounded distribution. In Section 5, we establish the uniform continuity of the entropies, a mathematical result that may be useful in further research.

2 Preliminaries

$\Sigma^* = \{0, 1\}^*$ is the set of all finite binary strings. The empty string is denoted by ϵ . Σ^n is the set of strings of length n and $|x|$ denotes the length of the string x . Strings are lexicographically ordered. The logarithm of x in base 2 is denoted by $\log(x)$. The real interval between a and b , including a and excluding b is represented by $[a, b)$. A sequence of real numbers r_n is denoted by $(r_n)_{n \in \mathbb{N}}$.

2.1 Kolmogorov Complexity

Kolmogorov complexity was introduced independently, in the 1960's by Solomonoff (7), Kolmogorov (8), and Chaitin (9). Only essential definitions and basic results are given here, for further details see (1). The model of computation used is the prefix-free Turing machine, *i.e.*, Turing machines with a prefix-free domain. A set of strings A is prefix-free if no string in A is prefix of another string of A . Kraft's inequality guarantees that for any prefix-free set A , $\sum_{x \in A} 2^{-|x|} \leq 1$. In this work all resource bounds t considered are time constructible, *i.e.*,

there is a Turing machine whose running time is exactly $t(n)$ on every input of size n .

Definition 2.1. Kolmogorov complexity. *Let U be a fixed prefix-free universal Turing machine. For any two strings $x, y \in \Sigma^*$, the Kolmogorov complexity of x given y is $K(x|y) = \min_p \{|p| : U(p, y) = x\}$, where $U(p, y)$ is the output of the program p with auxiliary input y when it is run in the machine U . For any time constructible t , the t -time-bounded Kolmogorov complexity of x given y is, $K^t(x|y) = \min_p \{|p| : U(p, y) = x \text{ in at most } t(|x|) \text{ steps}\}$.*

The default value for y , the auxiliary input is the empty string ϵ ; for simplicity, we denote $K(x|\epsilon)$ and $K^t(x|\epsilon)$ by $K(x)$ and $K^t(x)$, respectively. The choice of the universal Turing machine affects the running time of a program by, at most, a logarithmic factor and the program length by, at most, a constant number of extra bits.

Definition 2.2. *Let c be a non-negative integer. We say that $x \in \Sigma^n$ is c -Kolmogorov random if $K(x) \geq n - c$.*

Proposition 2.3. For all strings $x, y \in \Sigma^*$, we have:

- $K(x) \leq K^t(x) + O(1) \leq |x| + 2 \log(|x|) + O(1)$;
- $K(x|y) \leq K(x) + O(1)$ and $K^t(x|y) \leq K^t(x) + O(1)$;
- There are at least $2^n(1 - 2^{-c})$ c -Kolmogorov random strings of length n .

As Solovay (10) observed, for infinitely many x , the time-bounded version of Kolmogorov complexity equals the unbounded version. Formally, we have:

Theorem 2.4. (Solovay (10)) For all time-bounds $t(n) \geq n + O(1)$ we have $K^t(x) = K(x) + O(1)$ for infinitely many x .

As a consequence of this result, there is a string x of arbitrarily large Kolmogorov complexity such that $K^t(x) = K(x) + O(1)$.

Definition 2.5. A semi-measure over a discrete set X is a function $f : X \rightarrow [0, 1]$ such that, $\sum_{x \in X} f(x) \leq 1$. We say that a semi-measure is a measure if the equality holds. A semi-measure is constructive if it is semi-computable from below, i.e., for each x , there is a Turing machine that produces a monotone increasing sequence of rationals converging to $f(x)$.

An important constructive semi-measure based on Kolmogorov complexity is defined by $\mathbf{m}(x) = 2^{-K(x)}$. This semi-measure dominates any other constructive semi-measure μ (see (11; 12)), in the sense that there is a constant $c_\mu = 2^{-K(\mu)}$ such that, for all x , $\mathbf{m}(x) \geq c_\mu \mu(x)$. For this reason, this semi-measure is called universal. Since it is natural to consider time-bounds on Kolmogorov complexity, we can define $\mathbf{m}^t(x)$, a time-bounded version of $\mathbf{m}(x)$.

Definition 2.6. We say that a function f is computable in time t if there is a Turing machine that on the input x computes the output $f(x)$, in exactly $t(|x|)$ steps.

Definition 2.7. The t -time-bounded universal distribution is $\mathbf{m}^t(x) = c2^{-K^t(x)}$, where c is a real number such that $\sum_{x \in \Sigma^*} \mathbf{m}^t(x) = 1$.

In (1), the authors proved that $\mathbf{m}^{t'}$ dominates distributions computable in time t , where t' is a time-bound that only depends on t . Formally:

Theorem 2.8. (Claim 7.6.1 in (1)) If μ^* , the cumulative probability distribution of μ , is computable in time $t(n)$ then for all $x \in \Sigma^*$, $\mathbf{m}^{t'}(x) \geq 2^{-K^{t'}(\mu^*)} \mu(x)$, where $t'(n) = nt(n) \log(nt(n))$.

2.2 Entropies

Information Theory was introduced in 1948 by C.E. Shannon (2). Shannon entropy quantifies the uncertainty of the results of an experiment; it quantifies the average number of bits necessary to describe an outcome from an event.

Definition 2.9. (Shannon entropy (2)) Let \mathcal{X} be a finite or infinitely countable set and let X be a random variable taking values in \mathcal{X} with distribution P . The Shannon entropy of the random variable X is

$$H(X) = - \sum_{x \in \mathcal{X}} P(x) \log P(x) \quad (2)$$

Definition 2.10. (Rényi entropy (3)) Let \mathcal{X} be a finite or infinitely countable set and let X be a random variable taking values in \mathcal{X} with distribution P and let $\alpha \neq 1$ be a non-negative real number. The Rényi entropy of order α of the random variable X is

$$R_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{x \in \mathcal{X}} P(x)^\alpha \right) \quad (3)$$

Definition 2.11. (Tsallis entropy (4)) Let \mathcal{X} be a finite or infinitely countable set and let X be a random variable taking values in \mathcal{X} with distribution P and let $\alpha \neq 1$ be a non-negative real number. The Tsallis entropy of order α of the random variable X is

$$T_\alpha(X) = \frac{1}{\alpha-1} \left(1 - \sum_{x \in \mathcal{X}} P(x)^\alpha \right) \quad (4)$$

It is easy to prove that

$$\lim_{\alpha \rightarrow 1} R_\alpha(X) = \lim_{\alpha \rightarrow 1} T_\alpha(X) = H(X) \quad (5)$$

Given the conceptual differences in the definitions of Kolmogorov complexity and Shannon entropy, it is interesting that under some weak restrictions on the distribution of the strings, they are related, in the sense that the value of Shannon entropy equals the expected value of Kolmogorov complexity, up to a constant term that only depends on the distribution.

Theorem 2.12. *Let $P(x)$ be a recursive probability distribution such that $H(P) < \infty$. Then,*

$$0 \leq \sum_x P(x)K(x) - H(P) \leq K(P) \quad (6)$$

Proof. (Sketch, see (1) for details). The first inequality follows directly from the Noiseless Coding Theorem, that, for these distributions, states $H(P) \leq \sum_x P(x)K(x)$. Since \mathbf{m} is universal, $\mathbf{m}(x) \geq 2^{-K(P)}P(x)$, for all x , which is equivalent to $\log P(x) \leq K(P) - K(x)$. Thus, we have:

$$\sum_x P(x)K(x) - H(P) = \sum_x (P(x)(K(x) + \log P(x))) \quad (7)$$

$$\leq \sum_x (P(x)(K(x) + K(P) - K(x))) \quad (8)$$

$$= K(P) \quad (9)$$

□

3 Kolmogorov Complexity and Entropy: How Close?

Since Rényi and Tsallis entropies are generalizations of Shannon entropy, we now study if Theorem 2.12 can be generalized for these entropies. Then, we prove that for distributions such that the cumulative probability distribution is computable in time $t(n)$, Shannon entropy equals the expected value of the t -time-bounded Kolmogorov complexity.

First, we observe that the interval $[0, K(P)]$ of the inequalities of Theorem 2.12 is tight up to a constant term that only depends on the universal Turing machine chosen as reference. A similar study has been included in (1) and in (13). The following examples illustrate the tightness of this interval. We present a probability distribution that satisfies:

$$\sum_x P(x)K(x) - H(P) = K(P) - O(1), \text{ with } K(P) \approx n \quad (10)$$

and a probability distribution that satisfies:

$$\sum_x P(x)K(x) - H(P) = O(1) \text{ and } K(P) \approx n. \quad (11)$$

Example 3.1. Fix $x_0 \in \Sigma^n$. Consider the distribution concentrated in x_0 , i.e.,

$$P_n(x) = \begin{cases} 1 & \text{if } x = x_0 \\ 0 & \text{otherwise} \end{cases}$$

Notice that describing this distribution is equivalent to describing x_0 . So, $K(P_n) = K(x_0) + O(1)$. On the other hand, $\sum_x P_n(x)K(x) - H(P_n) = K(x_0)$. Thus, if x_0 is c -Kolmogorov random, i.e., $K(x_0) \geq n - c$, then $K(P_n) \approx n$ and $\sum_x P_n(x)K(x) - H(P_n) \approx n$.

Example 3.2. Let y be a string of length n that is c -Kolmogorov random, i.e., $K(y) \geq n - c$ and consider the following probability distribution over Σ^* :

$$P_n(x) = \begin{cases} 0.y & \text{if } x = x_0 \\ 1 - 0.y & \text{if } x = x_1 \\ 0 & \text{otherwise} \end{cases}$$

where $0.y$ represents the real number between 0 and 1 which binary representation is y . Notice that we can choose x_0 and x_1 such that $K(x_0) = K(x_1) \leq c'$, where c' is a constant greater than 1 and hence does not depend on n . Thus,

1. $K(P_n) \geq n - c$, since describing P_n is essentially equivalent to describe x_0 , x_1 and y ;
2. $\sum_x P_n(x)K(x) = (0.y)K(x_0) + (1 - 0.y)K(x_1) \leq 0.y \times c' + (1 - 0.y) \times c' = c'$;
3. $H(P_n) = -0.y \log 0.y - (1 - 0.y) \log(1 - 0.y) \leq 1$.

Thus,

$$0 \leq \sum_x P_n(x)K(x) - H(P_n) \leq c' \ll K(P_n) \quad (12)$$

and

$$K(P_n) \geq n - c \quad (13)$$

Now we address the question if an analogue of Theorem 2.12 holds for Rényi and Tsallis entropies. We show that the Shannon entropy is the only entropy that verifies simultaneously both inequalities of Theorem 2.12 and thus is the only one suitable to deal with information.

For every $\varepsilon > 0$, $0 < \varepsilon' < 1$, and any probability distribution P , with finite support, (see (5)), we have:

$$R_{1+\varepsilon}(P) \leq H(P) \leq R_{1-\varepsilon'}(P) \quad (14)$$

Thus,

1. For $\alpha \geq 1$, $0 \leq \sum_x P(x)K(x) - R_\alpha(P)$;
2. For $\alpha \leq 1$, $\sum_x P(x)K(x) - R_\alpha(P) \leq K(P)$.

It is known that for a given probability distribution with finite support, the Rényi and Tsallis entropies are monotonic increasing functions one of each other with respect to α (see (14)). Thus, for every $\varepsilon > 0$ and $0 < \varepsilon' < 1$, we also have a similar relationship for the Tsallis entropy, i.e.,

$$T_{1+\varepsilon}(P) \leq H(P) \leq T_{1-\varepsilon'}(P) \quad (15)$$

Hence, it follows that:

1. For $\alpha \geq 1$, $0 \leq \sum_x P(x)K(x) - T_\alpha(P)$;
2. For $\alpha \leq 1$, $\sum_x P(x)K(x) - T_\alpha(P) \leq K(P)$.

In the next result we show that the inequalities above are, in general, false for different values of α .

Proposition 3.3. *There are recursive probability distributions P such that:*

1. $\sum_x P(x)K(x) - R_\alpha(P) > K(P)$, where $\alpha > 1$;
2. $\sum_x P(x)K(x) - R_\alpha(P) < 0$, where $\alpha < 1$;
3. $\sum_x P(x)K(x) - T_\alpha(P) > K(P)$, where $\alpha > 1$;

4. $\sum_x P(x)K(x) - T_\alpha(P) < 0$, where $\alpha < 1$.

Proof. For $x \in \Sigma^n$, consider the following probability distribution:

$$P_n(x) = \begin{cases} 1/2 & \text{if } x = 0^n \\ 2^{-n} & \text{if } x = 1x', x' \in \{0, 1\}^{n-1} \\ 0 & \text{otherwise} \end{cases}$$

It is clear that this distribution is recursive. We use this distribution for some specific n to prove all items.

1. First observe that:

$$H(P_n) = - \sum_x P_n(x) \log P_n(x) \quad (16)$$

$$= - \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2^n} 2^{n-1} \log \frac{1}{2^n} \right) \quad (17)$$

$$= - \left(-\frac{1}{2} - \frac{1}{2^n} 2^{n-1} n \right) \quad (18)$$

$$= \frac{n+1}{2} \quad (19)$$

Notice also that to describe P_n it is sufficient to give n , so $K(P_n) \leq c \log n$, where c is a real number.

By Theorem 2.12, we have,

$$\sum_x P_n(x)K(x) - H(P_n) \geq 0 \quad (20)$$

which implies that:

$$\sum_x P_n(x)K(x) \geq \frac{n+1}{2} \quad (21)$$

On the other hand, by definition:

$$R_\alpha(P_n) = \frac{1}{1-\alpha} \log \sum_x P_n(x)^\alpha \quad (22)$$

$$= \frac{1}{1-\alpha} \log \left(\frac{1}{2^\alpha} + 2^{n-1} \times \frac{1}{2^{n\alpha}} \right) \quad (23)$$

$$= \frac{1}{1-\alpha} \left(\log(2^{(n-1)\alpha} + 2^{n-1}) - n\alpha \right) \quad (24)$$

To prove that $\sum_x P_n(x)K(x) - R_\alpha(P_n) > K(P_n)$, it is sufficiently to prove that:

$$\lim_{n \rightarrow \infty} \left(\sum_x P_n(x)K(x) - R_\alpha(P_n) - K(P_n) \right) > 0 \quad (25)$$

i.e., the limit of the following expression is bigger than 0:

$$\frac{n+1}{2} - \frac{1}{1-\alpha} \left(\log(2^{(n-1)\alpha} + 2^{n-1}) - n\alpha \right) - c \log n \quad (26)$$

But,

$$\lim_{n \rightarrow \infty} \left(\frac{n+1}{2} - \frac{\log(2^{(n-1)\alpha} + 2^{n-1})}{1-\alpha} + \frac{n\alpha}{1-\alpha} - c \log n \right) \quad (27)$$

$$\geq \lim_{n \rightarrow \infty} \left(\frac{n+1}{2} + \frac{\log(2^{(n-1)\alpha})}{\alpha-1} - \frac{n\alpha}{\alpha-1} - c \log n \right) \quad (28)$$

$$= \lim_{n \rightarrow \infty} \left(\frac{n+1}{2} - \frac{\alpha}{\alpha-1} - c \log n \right) = +\infty \quad (29)$$

2. To prove this item we use the other inequality of Theorem 2.12:

$$\sum_x P_n(x)K(x) - H(P_n) \leq K(P_n) \quad (30)$$

which implies that

$$\sum_x P_n(x)K(x) \leq \frac{n+1}{2} + c \log n \quad (31)$$

So,

$$\sum_x P_n(x)K(x) - R_\alpha(P_n) \leq \frac{n+1}{2} + c \log n - \frac{1}{1-\alpha} \left(\log(2^{(n-1)\alpha} + 2^{n-1}) - n\alpha \right) \quad (32)$$

$$\leq \frac{n+1}{2} + c \log n - \frac{\log(2^{n-1})}{1-\alpha} + \frac{n\alpha}{1-\alpha} \quad (33)$$

$$= -\frac{n}{2} + \frac{1}{2} + c \log n + \frac{1}{1-\alpha} \quad (34)$$

Thus, taking n sufficiently large, the conclusion follows.

3. The Tsallis entropy of order α of distribution P_n is

$$T_\alpha(P_n) = \frac{1}{\alpha-1} - \frac{1}{\alpha-1} \sum_x P_n(x)^\alpha \quad (35)$$

$$= \frac{1}{\alpha-1} - \frac{1}{\alpha-1} \left(\frac{1}{2^\alpha} + \frac{1}{2^{n\alpha}} \times 2^{n-1} \right) \quad (36)$$

$$= \frac{2^\alpha - 1}{2^\alpha(\alpha-1)} - \frac{2^{n(1-\alpha)}}{2(\alpha-1)} \quad (37)$$

Using the inequality 21, we get

$$\sum_x P_n(x)K(x) - T_\alpha(P_n) = \sum_x P_n(x)K(x) - \frac{2^\alpha - 1}{2^\alpha(\alpha-1)} + \frac{2^{n(1-\alpha)}}{2(\alpha-1)} \quad (38)$$

$$\geq \frac{n+1}{2} - \frac{2^\alpha - 1}{2^\alpha(\alpha-1)} + \frac{2^{n(1-\alpha)}}{2(\alpha-1)} \quad (39)$$

Since $\alpha > 1$, for n sufficiently large,

$$\frac{n+1}{2} - \frac{2^\alpha - 1}{2^\alpha(\alpha-1)} + \frac{2^{n(1-\alpha)}}{2(\alpha-1)} \geq \frac{n+1}{2} - \frac{2^\alpha - 1}{2^\alpha(\alpha-1)} \quad (40)$$

$$> c \log n \geq K(P_n) \quad (41)$$

4. Using the inequality 31, we get

$$c \sum_x P_n(x)K(x) - T_\alpha(P_n) \leq \frac{n+1}{2} + c \log n - \frac{2^\alpha - 1}{2^\alpha(\alpha-1)} + \frac{2^{n(1-\alpha)}}{2(\alpha-1)} \quad (42)$$

Since $\alpha < 1$, for n sufficiently large, we conclude that:

$$\frac{n+1}{2} + c \log n + \frac{2^\alpha - 1}{2^\alpha(1-\alpha)} - \frac{2^{n(1-\alpha)}}{2(1-\alpha)} < 0 \quad (43)$$

□

We generalize this result, obtaining the following theorem which proof is similar to the previous proposition.

Theorem 3.4. *For every $\Delta > 0$ and $\alpha > 1$ there are recursive probability distributions P such that,*

$$1. \sum_x P(x)K(x) - R_\alpha(P) \geq (K(P))^\alpha;$$

2. $\sum_x P(x)K(x) - R_\alpha(P) \geq K(P) + \Delta;$
3. $\sum_x P(x)K(x) - T_\alpha(P) \geq (K(P))^\alpha;$
4. $\sum_x P(x)K(x) - T_\alpha(P) \geq K(P) + \Delta.$

The previous results show that only the Shannon entropy satisfies the inequalities of Theorem 2.12. Now, we analyze if a similar relationship holds in a time-bounded Kolmogorov complexity scenario.

If, instead of considering $K(P)$ and $K(x)$ in the inequalities of Theorem 2.12, we use their time-bounded version and impose some computational restrictions on the distributions, we obtain a similar result. Notice that, for the class of distributions mentioned on the following theorem, the entropy equals (up to a constant) the expected value of time-bounded Kolmogorov complexity.

Theorem 3.5. *Let P be a probability distribution over Σ^n such that P^* , the cumulative probability distribution of P , is computable in time $t(n)$. Setting $t'(n) = O(nt(n) \log(nt(n)))$, we have,*

$$0 \leq \sum_x P(x)K^{t'}(x) - H(P) \leq K^{t'}(P^*) \quad (44)$$

Proof. The first inequality follows directly from the similar inequality of Theorem 2.12 and from the fact that $K^t(x) \geq K(x)$.

From Theorem 2.8, if P is a probability distribution such that P^* is computable in time $t(n)$, then for all $x \in \Sigma^n$ and $t'(n) = nt(n) \log(nt(n))$, $K^{t'}(x) + \log P(x) \leq K^{t'}(P^*)$. Then, summing over all x , we get:

$$\sum_x P(x)(K^{t'}(x) + \log P(x)) \leq \sum_x P(x)K^{t'}(P^*) \quad (45)$$

which is equivalent to $\sum_x P(x)K^{t'}(x) - H(P) \leq K^{t'}(P^*)$. \square

4 On the Entropies of the Time-Bounded Universal Distribution

If the time that a program can use to produce a string is bounded, we get the so called time-bounded universal distribution, $\mathbf{m}^t(x) = c2^{-K^t(x)}$. In this section, we study the convergence of the three entropies with respect to this distribution.

Theorem 4.1. *The Shannon entropy of the distribution \mathbf{m}^t diverges.*

Proof. If $x \geq 2$ then $f(x) = x2^{-x}$ is a decreasing function. Let A be the set of strings such that $-\log \mathbf{m}^t(x) \geq 2$; this set is recursively enumerable.

$$H(\mathbf{m}^t) = \sum_{x \in \Sigma^*} -\mathbf{m}^t(x) \log \mathbf{m}^t(x) \quad (46)$$

$$\geq \sum_{x \in A} -\mathbf{m}^t(x) \log \mathbf{m}^t(x) \quad (47)$$

$$= \sum_{x \in A} c2^{-K^t(x)} (K^t(x) - \log c) \quad (48)$$

$$= -c \log c \sum_{x \in A} 2^{-K^t(x)} + c \sum_{x \in A} K^t(x) 2^{-K^t(x)} \quad (49)$$

So, if we prove that $\sum_{x \in A} K^t(x) 2^{-K^t(x)}$ diverges, the result follows. Assume, by contradiction, that $\sum_{x \in A} K^t(x) 2^{-K^t(x)} < d$ for some $d \in \mathbb{R}$. Then, considering

$$r(x) = \begin{cases} \frac{1}{d} K^t(x) 2^{-K^t(x)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \quad (50)$$

we conclude that r is a semi-measure. Thus, there is a constant c' such that, for all x , $r(x) \leq c' \mathbf{m}(x)$. Hence, for $x \in A$, we have

$$\frac{1}{d} K^t(x) 2^{-K^t(x)} \leq c' 2^{-K(x)} \quad (51)$$

So, $K^t(x) \leq c' d 2^{K^t(x)-K(x)}$, which is a contradiction since by Theorem 2.4, A contains infinitely many strings x of time-bounded Kolmogorov complexity larger than a constant such that $K^t(x) = K(x) + O(1)$. The contradiction follows from the assumption that the sum $\sum_{x \in A} K^t(x) 2^{-K^t(x)}$ converges. So, $H(\mathbf{m}^t)$ diverges. \square

Now we show that, similarly to the behavior of Rényi and Tsallis entropies of universal distribution \mathbf{m} (see (15)), we have $R_\alpha(\mathbf{m}^t) < \infty$ iff $\alpha > 1$ and $T_\alpha(\mathbf{m}^t) < \infty$ iff $\alpha > 1$. First observe that:

1. If $\alpha > 1$, $T_\alpha(P) \leq \frac{1}{\alpha-1} + R_\alpha(P)$;
2. If $\alpha < 1$, $T_\alpha(P) \geq \frac{1}{\alpha-1} + R_\alpha(P)$.

Theorem 4.2. *The Tsallis entropy of order α of time-bounded universal distribution \mathbf{m}^t converges iff $\alpha > 1$.*

Proof. From Theorem 8 of (15), we have that $\sum_{x \in \Sigma^*} (\mathbf{m}(x))^\alpha$ converges if $\alpha > 1$. Since \mathbf{m}^t is a probability distribution, there is a constant λ such that, for all x , $\mathbf{m}^t(x) \leq \lambda \mathbf{m}(x)$. So, $(\mathbf{m}^t(x))^\alpha \leq (\lambda \mathbf{m}(x))^\alpha$, which implies that

$$\sum_{x \in \Sigma^*} (\mathbf{m}^t(x))^\alpha \leq \lambda^\alpha \sum_{x \in \Sigma^*} (\mathbf{m}(x))^\alpha \quad (52)$$

from which we conclude that for $\alpha > 1$, $T_\alpha(\mathbf{m}^t)$ converges.

For $\alpha < 1$, the proof is analogous to the proof of Theorem 4.1. Suppose that $\sum_{x \in \Sigma^*} (\mathbf{m}^t(x))^\alpha < d$ for some $d \in \mathbb{R}$. Hence, $r(x) = \frac{1}{d} (\mathbf{m}^t(x))^\alpha$ is a constructible semi-measure. Then, there is a constant τ such that for all $x \in \Sigma^*$,

$$r(x) = \frac{1}{d} (c 2^{-K^t(x)})^\alpha \leq \tau 2^{-K(x)} \quad (53)$$

which is equivalent to

$$\frac{c^\alpha}{d\tau} \leq 2^{\alpha K^t(x) - K(x)} \quad (54)$$

By Theorem 2.4, there are infinitely many strings x such that $K^t(x) = K(x) + O(1)$. Then it would follow that for these strings $\frac{c^\alpha}{d\tau} \leq 2^{(\alpha-1)K(x)}$, which is false since these particular strings can have arbitrarily large Kolmogorov complexity. \square

Theorem 4.3. *The Rényi entropy of order α of time-bounded universal distribution \mathbf{m}^t converges iff $\alpha > 1$.*

Proof. For $\alpha > 1$, since $\sum_x 2^{-K^t(x)} < +\infty$, we have $\sum_x (2^{-K^t(x)})^\alpha < \infty$. Thus, $R_\alpha(\mathbf{m}^t)$ converges.

For $\alpha < 1$, suppose without loss of generality that α is rational (otherwise take another rational slightly larger than α). Assume that $\sum_x (2^{-K^t(x)})^\alpha < \infty$.

Then by universality of \mathbf{m} and since $(2^{-K^t(x)})^\alpha$ is computable, we would have $\mathbf{m}(x) \geq (2^{-K^t(x)})^\alpha$ which, by taking logarithms, is equivalent to $K^t(x) \geq \frac{1}{\alpha} K(x) + O(1)$. Since $(1/\alpha) > 1$, this would contradict Solovay's Theorem of page 3. \square

5 Uniform Continuity of the Entropies

Shannon, Rényi, and Tsallis entropies are very useful measures in Information Theory and Physics. In this section we look to these entropies as functions of the corresponding probability distribution, and establish an important property: all the three entropies are uniformly continuous. We think that this mathematical result may be useful in the future research on this topic. It should be mentioned that the uniform continuity of the Tsallis entropy for certain values of α has already been proved in (16).

5.1 Tsallis Entropy

In order to prove the uniform continuity of the Tsallis entropy, we need some technical lemmas.

Lemma 5.1. *Let $0 \leq b < a \leq 1$ and $0 < \alpha < 1$. We have,*

$$(a - b)^\alpha \geq a^\alpha - b^\alpha \quad (55)$$

Proof. Consider the function $f(x) = x^\alpha$ for $0 < \alpha < 1$ and $0 \leq x \leq 1$. So, $f'(x) = \alpha x^{\alpha-1}$ is positive for all $0 \leq x \leq 1$ and $f''(x) = \alpha(\alpha - 1)x^{\alpha-2}$ is negative. We want to prove that:

$$f(a - b) > f(a) - f(b) \quad (56)$$

If we shift to the left a and b ,

$$\begin{cases} a' = a - e \\ b' = b - e \end{cases} \quad (57)$$

it is easy to show that the value of $f(a') - f(b')$ (for $e \geq 0$) is greater than $f(a) - f(b)$. In particular, if $e = b$,

$$f(a') - f(b') = f(a - b) - 0 = f(a - b) > f(a) - f(b) \quad (58)$$

which means that $(a - b)^\alpha > a^\alpha - b^\alpha$. \square

Lemma 5.2. *Let $0 \leq b < a \leq 1$, and $\alpha > 1$. We have,*

$$a^\alpha - b^\alpha \leq \alpha(a - b) \quad (59)$$

Proof. Consider the function $f(x) = x^\alpha$ for $\alpha > 1$ and $0 \leq x \leq 1$. Let $0 \leq a \leq 1$, $0 \leq b \leq 1$. The function f is continuous in $[a, b]$ and differentiable in (a, b) . So, by Lagrange's Theorem,

$$\exists c \in (a, b) : a^\alpha - b^\alpha = f'(c)(a - b) \quad (60)$$

Note that $f'(c) = \alpha c^{\alpha-1} < \alpha$. Thus, $c(a - b) < \alpha(a - b)$. \square

Theorem 5.3. *Let P and Q be two probability distributions over Σ^n . Let $T_\alpha(P)$ and $T_\alpha(Q)$ be the Tsallis entropy of distributions P and Q , respectively, where $\alpha > 0$, $\alpha \neq 1$. Then,*

$$\forall \varepsilon > 0, \exists \delta > 0, \forall P, Q : \max_x |P(x) - Q(x)| \leq \delta \Rightarrow |T_\alpha(P) - T_\alpha(Q)| \leq \varepsilon \quad (61)$$

Proof. If $\alpha < 1$, we have:

$$|T_\alpha(P) - T_\alpha(Q)| = \left| \frac{1}{\alpha - 1} + \frac{1}{1 - \alpha} \sum_x P(x)^\alpha - \frac{1}{\alpha - 1} - \frac{1}{1 - \alpha} \sum_x Q(x)^\alpha \right| \quad (62)$$

$$= \frac{1}{1 - \alpha} \left| \sum_x P(x)^\alpha - \sum_x Q(x)^\alpha \right| \quad (63)$$

$$= \frac{1}{1 - \alpha} \left| \sum_x (P(x)^\alpha - Q(x)^\alpha) \right| \quad (64)$$

$$\leq \frac{1}{1 - \alpha} \sum_x |P(x) - Q(x)|^\alpha, \text{ by Lemma 5.1} \quad (65)$$

$$\leq \frac{1}{1 - \alpha} \sum_x \delta^\alpha = \frac{2^n \times \delta^\alpha}{1 - \alpha} \quad (66)$$

Thus, it is sufficient to consider $\delta = \sqrt[\alpha]{\frac{(1-\alpha)\varepsilon}{2^n}}$.

If $\alpha > 1$, we have:

$$|T_\alpha(P) - T_\alpha(Q)| = \left| \frac{1}{\alpha-1} + \frac{1}{1-\alpha} \sum_x P(x)^\alpha - \frac{1}{\alpha-1} - \frac{1}{1-\alpha} \sum_x Q(x)^\alpha \right| \quad (67)$$

$$= \frac{1}{\alpha-1} \left| \sum_x P(x)^\alpha - \sum_x Q(x)^\alpha \right| \quad (68)$$

$$= \frac{1}{\alpha-1} \left| \sum_x (P(x)^\alpha - Q(x)^\alpha) \right| \quad (69)$$

$$\leq \frac{1}{\alpha-1} \sum_x \alpha |P(x) - Q(x)|, \text{ by Lemma 5.2} \quad (70)$$

$$\leq \frac{1}{\alpha-1} \sum_x \alpha \delta = \frac{\alpha 2^n}{\alpha-1} \delta \quad (71)$$

Thus, it is sufficient to consider $\delta = \frac{(\alpha-1)}{\alpha 2^n} \varepsilon$. \square

5.2 Shannon Entropy

Lemma 5.4. *The function $x \log x$ is uniformly continuous in $[0, 1]$.*

Proof. The function is continuous in $[0, 1]$ and this is a compact set, so the function is uniformly continuous. \square

Theorem 5.5. *Let P and Q be two probability distributions over Σ^n . Let $H(P)$ and $H(Q)$ be the Shannon entropy of distributions P and Q , respectively. Then,*

$$\forall \varepsilon > 0, \exists \delta > 0, \forall P, Q : \max_x |P(x) - Q(x)| \leq \delta \Rightarrow |H(P) - H(Q)| \leq \varepsilon \quad (72)$$

Proof. By Lemma 5.4, we have that:

$$\forall \gamma > 0, \exists \beta > 0, \forall x, y : \max |x - y| \leq \beta \Rightarrow |x \log x - y \log y| \leq \gamma \quad (73)$$

So,

$$|H(P) - H(Q)| = \left| - \sum_x P(x) \log P(x) + \sum_x Q(x) \log Q(x) \right| \quad (74)$$

$$= \left| \sum_x (P(x) \log P(x) - Q(x) \log Q(x)) \right| \quad (75)$$

$$\leq \sum_x |P(x) \log P(x) - Q(x) \log Q(x)| \quad (76)$$

$$\leq \sum_x \gamma = 2^n \gamma \quad (77)$$

It is sufficient to consider $\gamma = \frac{\varepsilon}{2^n}$ and consider $\delta = \beta$. \square

5.3 Rényi Entropy

Lemma 5.6. *Let a and b be two real numbers greater than 1. Then,*

$$|\log a - \log b| \leq 2|a - b| \quad (78)$$

Proof. Consider the function $f(x) = \log x$ and $x \geq 1$. Let $a, b \geq 1$. The function f is continuous in $[a, b]$ and differentiable in (a, b) . So, by Lagrange's Theorem,

$$\exists c \in (a, b) : |\log a - \log b| = f'(c)|a - b| \quad (79)$$

Note that $f'(c) = \frac{1}{\ln 2c} < 2$. Thus, $c|a - b| < 2|a - b|$. \square

Lemma 5.7. *The function x^α is uniformly continuous in $[0, 1]$. Thus,*

$$\forall \gamma > 0 \exists \beta > 0 \forall x, y : \max |x - y| \leq \beta \Rightarrow |x^\alpha - y^\alpha| \leq \gamma \quad (80)$$

Proof. The function is continuous in $[0, 1]$ and this is a compact set, so the function is uniformly continuous. \square

Theorem 5.8. *Let P and Q be two probability distributions over Σ^n . Let $R_\alpha(P)$ and $R_\alpha(Q)$ be the Rényi entropy of distributions P and Q respectively, where $\alpha > 0$, $\alpha \neq 1$. Then,*

$$\forall \varepsilon > 0, \exists \delta > 0, \forall P, Q : \max_x |P(x) - Q(x)| \leq \delta \Rightarrow |R_\alpha(P) - R_\alpha(Q)| \leq \varepsilon \quad (81)$$

Proof. If $\alpha < 1$:

Consider $\gamma = \frac{\varepsilon(1-\alpha)}{2^{n+1}}$. By Lemma 5.7, we know that there is β such that if $\max |x - y| \leq \beta$ then $|x^\alpha - y^\alpha| \leq \gamma$. So, consider $\delta = \beta$.

We have to show that if $\max_x |P(x) - Q(x)| \leq \delta$ then $|R_\alpha(P) - R_\alpha(Q)| \leq \varepsilon$:

$$|R_\alpha(P) - R_\alpha(Q)| = \left| \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha - \frac{1}{1-\alpha} \log \sum_x Q(x)^\alpha \right| \quad (82)$$

$$= \frac{1}{1-\alpha} \cdot \left| \log \sum_x P(x)^\alpha - \log \sum_x Q(x)^\alpha \right| \quad (83)$$

$$\leq \frac{2}{1-\alpha} \left| \sum_x (P(x)^\alpha - Q(x)^\alpha) \right|, \text{ by Lemma 5.6} \quad (84)$$

$$\leq \frac{2}{1-\alpha} \sum_x |P(x)^\alpha - Q(x)^\alpha| \quad (85)$$

$$\leq \frac{2}{1-\alpha} \sum_x \frac{\varepsilon(1-\alpha)}{2^{n+1}}, \text{ by Lemma 5.7} \quad (86)$$

$$\leq \frac{2^{n+1}}{1-\alpha} \cdot \frac{\varepsilon(1-\alpha)}{2^{n+1}} = \varepsilon \quad (87)$$

If $\alpha > 1$:

One of the $P(x)$ is at least $1/2^n$ so that $\sum_x P(x)^\alpha \geq 1/2^{\alpha n}$ and $\frac{d}{dx} \log x = \frac{1}{x \ln 2} \leq \frac{2^{\alpha n}}{\ln 2}$; assume that $\max_x |P(x) - Q(x)| \leq \delta$. We have

$$\left| \log \sum_x P(x)^\alpha - \log \sum_x Q(x)^\alpha \right| \leq \frac{2^{\alpha n}}{\ln 2} \left| \sum_x P(x)^\alpha - \sum_x Q(x)^\alpha \right| \quad (88)$$

But, by Lemma 5.2 we have

$$\left| \sum_x P(x)^\alpha - \sum_x Q(x)^\alpha \right| \leq \sum_x |P(x)^\alpha - Q(x)^\alpha| \leq 2^n \delta \alpha \quad (89)$$

Thus,

$$\left| \log \sum_x P(x)^\alpha - \log \sum_x Q(x)^\alpha \right| \leq \frac{\alpha 2^{n(\alpha+1)} \delta}{\ln 2} \quad (90)$$

We get

$$|R_\alpha(p) - R_\alpha(q)| \leq \frac{\alpha 2^{n(\alpha+1)} \delta}{\ln 2(\alpha - 1)} \quad (91)$$

It is sufficient to consider $\delta = \frac{\ln 2(\alpha - 1)}{\alpha 2^{n(\alpha+1)}} \varepsilon$. □

6 Conclusions

We proved that, among the three entropies we have studied, Shannon entropy is the only one that satisfies the relationship with the expected value of Kolmogorov complexity stated in (1), by exhibiting a probability distribution for which the relationship fails for some values of α of Tsallis and Rényi entropies. Furthermore, under the assumption that cumulative probability distribution is computable in an allotted time, a time-bounded version of the same relationship holds for the Shannon entropy. Since it is natural to define a probability distribution based on time-bounded Kolmogorov complexity, we studied the convergence of this distribution under Shannon entropy and its two generalizations: Tsallis and Rényi entropies. We also proved that the three entropies considered in this work are uniformly continuous.

Acknowledgements

This work was partially supported by the project *CSI*² (PTDC/EIA- CCO/099951/2008), the grants SFRH/BD/33234/2007 and SFRH/BD/28419/2006 from Fundação para a Ciência e Tecnologia (FCT) and funds granted to Instituto de Telecomunicações and Laboratório de Inteligência Artificial e Ciência de Computadores through the Programa de Financiamento Plurianual.

References

1. Li, M.; Vitányi, P. An Introduction to Kolmogorov Complexity and Its Applications, 3rd ed.; Springer Publishing Company, Incorporated: New York, NY, USA, 2008.
2. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
3. Rényi, A. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Statistical Laboratory of the University of California, Berkeley, CA, USA, 20 June–30 July, 1960; pp. 547–561.
4. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
5. Cachin, C. Entropy measures and unconditional security in cryptography, 1st ed.; In *ETH Series in Information Security and Cryptography*; Hartung-Gorre Verlag: Konstanz, Germany, 1997.
6. Cover, T.; Gacs, P.; Gray, M. Kolmogorov's contributions to information theory and algorithmic complexity. *Ann. Probab.* **1989**, *17*, 840–865.
7. Solomonoff, R. A formal theory of inductive inference; Part I. *Inform. Contr.* **1964**, *7*, 1–22.
8. Kolmogorov, A.N. Three approaches to the quantitative definition of information. *Probl. Inform. Transm.* **1965**, *1*, 1–17.

9. Chaitin, G.J. On the length of programs for computing finite binary sequences. *J. ACM* **1966**, *13*, 547–569.
10. Solovay, R.M. Draft of a paper (or series of papers) on Chaitin’s work. IBM Thomas J. Watson Research Center, LOCATION. Unpublished manuscript, 1975.
11. Levin, L. Laws of information conservation (non-growth) and aspects of the foundation of probability theory. *Prob. Peredachi Inf.* **1974**, *10*, 206–210.
12. Gacs, P. On the symmetry of algorithmic information. *Soviet Mathematics Doklady* **1974**, *15*, 1477–1480.
13. Grunwald, P.; Vitanyi, P. Shannon information and Kolmogorov complexity. 2004, arXiv:cs/0410002v1.
14. Ramshaw, J. Thermodynamic stability conditions for the Tsallis and Rényi entropies. *Phys. Lett. A* **1995**, *198*, 119 – 121.
15. Tadaki, K. The Tsallis entropy and the Shannon entropy of a universal probability. In Proceedings of the International Symposium on Information Theory, Toronto, Canada, 6–11 July 2008; abs/0805.0154, pp. 2111–2115.
16. Zhang, Z. Uniform estimates on the Tsallis entropies. *Lett. Math. Phys.* **2007**, *80*, 171–181.